

STIC-Biotech/ChemLib

71965

From: Helmer, Georgia
Sent: Monday, July 29, 2002 10:59 AM
To: STIC-Biotech/ChemLib
Subject: 09/643,755

Could you please do sequence searches of SEQ ID 1, 2, and 3 of this case, 09/643,755.

SEQ ID 1 and 3 are DNA. SEQ ID 2 is protein.

Could you also do an oligo search of SEQ ID 1 and 3.

The commercial and inhouse databases.

Thanks in advance for you assistance.

Cheymsen

Georgia L. Helmer Ph.D.
Patent Examiner
Crystal Mall 1, 9D14
AU 1638
703-308-7023
Georgia.Helmer@USPTO.gov
mailbox 9e12

SPRMS *61.*

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

01/13/03
591

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 7/31
Date Completed: 8/2
Searcher Prep/Review: 12
Clerical: _____
Online time: 12

TYPE OF SEARCH:
NA Sequences: 4
AA Sequences: 1
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 48
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:44:09 ; Search time 9237.8 Seconds
(without alignments)
2657.216 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173

Sequence: 1 atgaattccttaagctctt.....gctagctaaagaatctga 1173

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: *
1: gb_da: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match Length	DB ID	Description
------------	-------	-------	--------------	-------	-------------

1	1173	100.0	1173	6	AX088019	AX088019 Sequence
2	1173	100.0	3957	6	AX088021	AX088021 Sequence
3	900.4	76.8	1305	4	BOYCHMOB	J00003 bovine chym
4	897.2	76.5	1240	6	AR002347	AR002347 Sequence
5	897.2	76.5	1460	6	E00042	E00042 DNA coding
6	896.2	76.4	1291	6	A15836	A15836 chymosin ge
7	895.6	76.4	1175	6	I04058	I04058 Sequence 5
8	894.8	76.3	2733	6	AR073077	AR073077 Sequence
9	894	76.2	1275	4	BOYCHMOA	J00002 bovine chym
10	894	76.2	1290	6	E00075	E00075 DNA encodi
11	894	76.2	1291	6	A15633	A15633 preprochymo
12	893	76.1	1289	6	E00144	E00144 DNA encodi
13	893	76.1	2726	6	I08097	I08097 Sequence 5
14	892.4	76.1	1311	6	E00108	E00108 DNA coding
15	892.4	76.1	1460	6	E00295	E00295 DNA encodi
16	890.8	75.9	1269	4	BOYCHMOC	J00004 bovine chym
17	886.8	75.6	1098	6	E00079	E00079 CDNA sequen
18	886.8	75.6	1098	6	E02341	E02341 CDNA sequen
19	872.8	74.4	2982	6	I08098	I08098 Sequence 1
20	870.8	74.2	1101	4	AF177290	AF177290 Bubalus a
21	839.6	71.6	1292	4	OAPPCY	X53037 Lamb mRNA f
22	776	66.2	1115	12	SYNPROCA	M22593 Synthetic b
23	753.6	64.2	1094	4	BP019786	U19786 Bos primige
24	749	63.9	1172	4	CDR131677	AJ131677 Camelus d
25	732.4	62.4	1230	4	SSU14406	U14406 Sus scrofa
26	718	61.2	1262	9	AB038386	AB038386 Callithr
27	627.2	53.5	969	6	A02000	A02000 Artificial
28	627.2	53.5	969	6	A06446	A06446 Artificial
29	619.6	52.8	1146	12	SYNPROCA	M31016 Synthetic b
30	605	51.6	1259	10	RNO251688	AJ251688 Rattus no
31	534	45.5	1140	6	AX19837	AX19837 DNA sequen
32	382.4	32.6	637	6	E05472	E05472 DNA sequen
33	381.8	32.5	1403	4	AB047246	AB047246 Canis fam
34	378.6	32.3	1161	4	AB047245	AB047245 Rhinoloph
35	373.8	31.9	1351	9	AB038384	AB038384 Callithr
36	370.8	31.6	1381	4	AB047244	AB047244 Sorex ung
37	363.8	31.0	1362	4	AB047243	AB047243 Suncus mu
38	363.8	31.0	1363	4	PIGPEPA	J04601 Pig pepsi
39	363.4	31.0	1221	4	CDR131678	AJ131678 Camelus d
40	362.2	30.9	1276	5	CHKPSN	D00215 Gallus gall
41	360.8	30.8	1266	4	RABPEP123	M59235 Rabbit peps
42	357.4	30.5	1353	9	MPEPEPA23	X59755 M. fuscata m
43	357.4	30.5	1361	9	MPEPEPA4	X59753 M. fuscata m
44	353.6	30.1	1371	4	RABPEP111	M59237 Rabbit peps
45	352.6	30.1	1338	9	MACPCA	M20788 Rhesus monk

ALIGNMENTS

RESULT 1
AX088019 LOCUS AX088019 1173 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 1 from Patent WO0114571.
ACCESSION AX088019
VERSION AX088019.1 GI:13396947
KEYWORDS
SOURCE
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 1173)
van Rooijen, G., Keon, R. G., Boothe, J. and Shen, Y.
TITLE Commercial Production of Chymosin in Plants
JOURNAL Patent: WO 0114571-A 1 01-MAR-2001;
Sembiosys Genetics Inc. (CA)
FEATURES
Location/Qualifiers
1..1173
/organism="Bos taurus"
/db_xref="taxon:9913"
1..1173
/note="unnamed protein product"

CDS

```
/codon_start=1
/protein_id="CAC34680.1"
/db_xref="GI:13396948"
/translation="MNFLLKSPFFAFLCFQGYFYAVTHAAETIRPLPKGSLKALK
EHLLEDFLOKQOYGISKSYSGFEVAVPLFNTNLDISOYFEKTIIGTPPOEFTVLEDT
GSSDFWVSTICKSNACKNHQRFDRKSTFONLGRPLSHYGTGSMGILIGYDITVY
SNVIDIOQTWGLSTQEPGDVFTVAEFDILMAVPSLASEYSEIFEDMMNRHLVAOD
LFSYMDRNGOESMLTLGAIDPSTYTGSLHWPVYVOOQWQFTVDSVTISGVVACEG
GCOALIDGTGSKLVGPSSDILIOQAIATNOYGEFDIDNDNLSYMPYVFEINGKM
YPLPMSAVTSDDOFCSTSGFSENHSOKWILGDVFIIRIYYSVPRANLVGLAKAI"
BASE COUNT      299 a      308 c      262 g      304 t
ORIGIN

Query Match      100.0%; Score 1173; DB 6; Length 1173;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaacttccttaagtccttcccttctcaagccttcccttggttcggtcaacttcgt 60
    |||
Db 1 ATGAACTTCCTTAAGTCCTTCCCTTCTACGCTTCCCTTGGTGGTCAATACCTTCGT 60

QY 61 gctgttactcaagcgtcgtgagatcaaccgcatctcctctacaaggtaagtcctcgt 120
    |||
Db 61 GCTGTTACTCAAGCTGCTGAGATCACCCGCAATTCCTCTACAAAGGTAACTCTCCGT 120

QY 121 aagcgctgaagaaacatggaactcttagaagaactcttgcagaaaacaagaatgagc 180
    |||
Db 121 AAGCGCTGAAGAAACATGGAACCTTCTAGAACTCTTGCAGAAACACAGATAGGATC 180

QY 181 agcagaactactccggttcgtgtaagttcgtagcgtgacacttacaacacttcgt 240
    |||
Db 181 AGCAGAACTACTCCGGCTTCGGTGAAGTGTCTAGCGTCCACTTACCAACTACCTTGA 240

QY 241 agtcaatacttgggaaagatctactcgtgaaccccgctcaagaagttaacggtctc 300
    |||
Db 241 AGTCAATACCTTGGGAAAGATCTACCTCGAAACCCGCTCAAGAGTTCAACCTTCTCT 300

QY 301 gatactggttcctcgtactctgtggttccctctactactgaagaagcaatgcttga 360
    |||
Db 301 GATACTGGTCTCCTGACTCTGCGGTTCCTCTATCTACTGCAAGAGCAATGCTCTGAAG 360

QY 361 aacaccaaagatctgactcgcagaagaagtcgacacttccagaacttaggcgaacct 420
    |||
Db 361 AACACCAAAGATTCGATCCGAGAAAGTCGTCACCTTCCAGAACTTAGGCAAACTTG 420

QY 421 tctataactcgtgacaggttagcatgaagaactcttagatgatacgttcactg 480
    |||
Db 421 TCTATACTAGTACGGTACAGGTAGCATGCAAGGAATCTTAGGCTATGATACCGTCACTG 480

QY 481 tccaacattgtgacatccaacagacagtagacttagaccaccaagaacaggtgatg 540
    |||
Db 481 TCCAACATTGTGACATTCACAGACAGTAGACTTAGACCCCAAGAACAGAGTGATGTC 540

QY 541 ttcacctatgcgaattcgtatgcatccttgatgcataccatcgctcgctcgaag 600
    |||
Db 541 TTCACCTATGCGAATTCGATGATGCTCTGATAGGCAATCCATCGCTCGCTCAGAG 600

QY 601 tactgataaccgtgttggacaacatgtagaacacacactgtgtgcttaagaactgttc 660
    |||
Db 601 TACTGATACCGTGTGTTGACACATGATGAAACGACACTTAGCTCAAGACTTGTTTC 660

QY 661 tcggtttacatgagacgaatgagccagagacagtcgtcaacgcttggagcttatcca 720
    |||
Db 661 TCGGTTTACATGAGACGAATGAGCAGAGACATGCTACGCTTGGAGCTATTGATCCA 720

QY 721 tctacttaacagagatcctctcaactgggttccagtaactgtgacagactggaatc 780
    |||
Db 721 TCTACTTAACAGAGATCTCTTCACTGAGGTTCACAGTCACTGACAGACTGCAATTC 780

QY 781 actgtgagacagtgtacacatcaagcgtgtgtgtgtgcatggaaggtgaatgtaact 840
    |||
Db 781 ACTGTGACAGTGTACACATCAAGCGGTGTGTGTGTGTCATGGAAGGTGATGTCAAGCT 840
```

```
QY 841 atcttggataccggtgactgccaaagctgtgctgcgaacttaagcagcacaattctcaattcag 900
    |||
Db 841 ATCTTGATACCGGATGACGTCACAAAGCTGTGCGGACTAGACAGGACATCTTCAACATTCAG 900

QY 901 caagcatatggagccacacagaacagtagtcggttgaattgatatagatgacacaactt 960
    |||
Db 901 CAAGCATATGGAGCCACACAGAACAGTAGTCGCTGAGTTGACATGATGTCACAACTT 960

QY 961 aactcatgctcactaggttcttcttgagatacaacgcaagatgtlaccatgaccccttc 1020
    |||
Db 961 AACTCATGCTCCTAGGTTCTCTTGAAGATCAACGCGCAAGATGTACCCTGACCCCTTC 1020

QY 1021 gccctacagcagcaagatgaaggttcgcacaggtgattcagaagtgaacattcc 1080
    |||
Db 1021 GCTTATACCGCAGCAGATGAAGGTTTCTGACCACTGATTCAGAGTGAACATTC 1080

QY 1081 cagaatgagatcttggagatgtgttcattcgttgagtaactacaagcgtcttgacagggc 1140
    |||
Db 1081 CAGAATGAGATCTTGGAGATGTGTCAATTCGTGAGTACTACAGCGTCTTGACAGGGCC 1140

QY 1141 aacaactcgttgggttagctaaagcaatcga 1173
    |||
Db 1141 AACAACTCGTTGGGCTACTAAGCAATCTGA 1173

RESULT 2
LOCUS AX088021 3957 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 3 from Patent WO0114571.
ACCESSION AX088021
VERSION AX088021.1 GI:13396949
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
REFERENCE
AUTHORS Van Rooijen, G., Keon, R. G., Boothe, J. and Shen, Y.
TITLE Commercial production of chymosin in plants
JOURNAL Patent: WO 0114571-A 3 01-MAR-2001;
Sembiosys Genetics Inc. (CA)
FEATURES
source
1..3957
/organism="synthetic construct"
/db_xref="taxon:32630"
1554..2726
/note="Figure 2"
/codon_start=1
/transl_table=11
/protein_id="CAC34680.1"
/db_xref="GI:13396950"
/translation="MNFLLKSPFFAFLCFQGYFYAVTHAAETIRPLPKGSLKALK
EHLLEDFLOKQOYGISKSYSGFEVAVPLFNTNLDISOYFEKTIIGTPPOEFTVLEDT
GSSDFWVSTICKSNACKNHQRFDRKSTFONLGRPLSHYGTGSMGILIGYDITVY
SNVIDIOQTWGLSTQEPGDVFTVAEFDILMAVPSLASEYSEIFEDMMNRHLVAOD
LFSYMDRNGOESMLTLGAIDPSTYTGSLHWPVYVOOQWQFTVDSVTISGVVACEG
GCOALIDGTGSKLVGPSSDILIOQAIATNOYGEFDIDNDNLSYMPYVFEINGKM
YPLPMSAVTSDDOFCSTSGFSENHSOKWILGDVFIIRIYYSVPRANLVGLAKAI"
BASE COUNT 1263 a 790 c 609 g 1295 t
ORIGIN

Query Match      100.0%; Score 1173; DB 6; Length 3957;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaacttccttaagtccttcccttctcaagccttcccttggttcggtcaacttcgt 60
    |||
Db 1554 ATGAACTTCCTTAAGTCCTTCCCTTCTACGCTTCCCTTGGTGGTCAATACCTTCGT 1613

QY 61 gctgttactcaagcgtcgtgagatcaaccgcatctcctctacaaggtaagtcctcgt 120
    |||
Db 1614 GCTGTTACTCAAGCTGCTGAGATCACCCGCAATTCCTCTACAAAGGTAACTCTCCGT 1673
```


QY 121 aaggcgctgaaggaacatgagcttctagaagacttcttcgagaacaacagatgacac 180
|||||
Db 1674 AAGGGGGAAGGAACAATGAGCTTCTAGAACACTTCTTGAGAAACAACAGATATGCAATC 1733
QY 181 agcagcaagtactccgcgcttcggtgaagtctgtagcgtgcacttaccactacattgat 240
|||||
Db 1734 AGCAGCAAGTACTCCGGCTTGGTGAAGTGTCTAGCCGTGACACTTACCACTACTCTGAT 1793
QY 241 agtaataactttggagaagatctacctcgaaaccccgctcaagagttaccggtctctt 300
|||||
Db 1794 AGTAATAACTTTGGAGAAGATCTACCTCGGAACCCCGCTCAAGAGTTTCAACCGTTCTCTT 1853
QY 301 gatactgctcctctgacttcgtggttccctctatctactcaagaagaatgcgctgcaag 360
|||||
Db 1654 GATACGTGTTCTCTGACTTCTGGTTCCCTCTATCTACTCAAGACGAATGCTCGCAAG 1913
QY 361 aaccacaagaatccgatccgagaagaagtcgacaccttcagaacttaggcaaaccttg 420
|||||
Db 1914 AACCAACCAAGATTCGATCCGAGAAAGCTCCACACTTCCAGAACTTAGCAAAACCCCTTG 1973
QY 421 tctatacactcaggttagcaggttagcagaagaatcttagctatgataccgctacgtctc 480
|||||
Db 1974 TCTTATACACTACGCTACAGGTACAGTACAGGAATCTTAGGCTATGATACCTGCACTGTC 2033
QY 481 tccaacattgtgacatccaacagacagtagacttagcaaccaagaacagcgtgagtctc 540
|||||
Db 2034 TCCAAACATTTGGAGCATTTCAACAGACAGTAGACTTAGACCAACAAGACGAGTATTC 2093
QY 541 ttcacactatgcaagaatccgatgacacacctgtgtagtgcataccacgcgtgcgacag 600
|||||
Db 2094 TTTCACCTATGCAAGAAATTCGATGCGATCCTTGATGAGCATACCATCTCGCTGCAAG 2153
QY 601 taccgagatccgtgttgaacaatgatacagcaaccctatgactcgaagacttgctc 660
|||||
Db 2154 TACTCGATACCTGTGTTTGAACATGATGAGACCGACACTTAGTACCTCAAGACTGTTTC 2213
QY 661 tcggtttacatgagcaaggaatgagcagagagacatgctcaacgcttgagactatgatcca 720
|||||
Db 2214 TCGGTTTACATGAGCAAGGAATGGCCAGAGAGCATGCTCAAGCTTGAATGATGCA 2273
QY 721 tctactacaacagagatctctcaactggttccagtcacgtgtagcagcagtagctgcaattc 780
|||||
Db 2274 TCCACACACACAGAGATCTTCTCACTGCGTTCCACTGCTGACAGCACTACTGGCAATTC 2333
QY 781 acgttgacaggtgcacacatcagcgtgtgtgttgcattgtagagtgagtgacagct 840
|||||
Db 2334 ACTGTGACAGTGTCCACCATCAGCGGTGTGTGTGATGTGAAGGTGGAATGTCAAGCT 2393
QY 841 atcttgataccggttagcgtccaagctggtcgagactagcagcagacatctcaacattcag 900
|||||
Db 2394 ATCTTGGATACCGGTAGCTCCAAAGCTGCTGCGACCTAGCAACGACATTTCTCAATTCAG 2453
QY 901 caagctattggagccacacagaaacagtaecggtgagattgacatagattgcgaacactt 960
|||||
Db 2454 CAACCTTTTGGAGCCACACAGAACACAGTACCGTGTGACATAGATTGCGACAACCTT 2513
QY 961 agctacatgctctacagttgtcttttggatcaagcaggaatgataccacttgaacccctcc 1020
|||||
Db 2514 AGCTACATGCTTACAGTGTCTTTGAGATCAACCGCAAGATGTACCACTGATACCCCTCC 2573
QY 1021 gctataccagcaggaatcaagggcttcgcaacagtgatctcagagtgagaacattcc 1080
|||||
Db 2574 GCTATATACAGCCAGATCAAGGGTCTTGCAACAGTGGATTCCAGAGTGAACAACATTC 2633
QY 1081 cagagaatgagatcttgagagatggttcatctcgttgagtactacagcgtctcttgacagggcc 1140
|||||
Db 2634 CAGAAATGGATCTTGGAGAGTGTCTTATTTGTGATGATACAGCTCTTTGACAGGGGCC 2693
QY 1141 aacaacctgltgggctagactaaagcaatctga 1173
|||||
Db 2694 AACCACTCTGTTGGGCTTAGCTAAAGCAATCTGA 2726

RESULT 3
BOVCHYMOB
LOCUS BOVCHYMOB 1305 bp mRNA linear MAM 26-APR-1993
DEFINITION bovine chymosin b (rennin) mRNA.
ACCESSION J00003
VERSION J00003.1 GI:162859
KEYWORDS chymosin; chymosin B; protease; rennin.
SOURCE bovine (calf) cdna of fourth stomach mucosa mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Harris,T.J.R., Lowe,P.A., Lyons,A., Thomas,P.G., Eaton,M.A.W.,
Mulligan,T.A., Patel,T.P., Bose,C.C., Carey,N.H., and Doel,M.T.
TITLE Molecular cloning and nucleotide sequence of cdna coding for calf
preprochymosin
JOURNAL Nucleic Acids Res. 10, 2177-2187 (1982)
MEDLINE 82221400
COMMENT
chymosin is the major proteolytic enzyme in the fourth stomach of
the unweaned calf. two chromatographically different forms, a and
b, of the enzyme and its precursor are known and a third form seems
likely (see bovchymoa, bovchymoc). this sequence has tentatively
been identified as coding for preprochymosin b. sequence comparison
indicates that the precursors for chymosins a and b differ by only
two amino acids, and for b and c by only four amino acids.
FEATURES
source
1..1305
/organism="Bos taurus"
/db_xref="taxon:9913"
<1..1305
/product="chymob mRNA"
26..1171
/note="preprochymosin b"
/codon_start=1
/protein_id="AAA30448.1"
/db_xref="GI:162860"
/translation="MRCLVLLVFLVSLGSAELTRPLVYKSKSLRKLKHGLEPDL
OKOQXGSSKSYSGFEVAPLIRNLDISOFGKRIYLICTPPQETVFLPDTSSPWPVS
ITKSNACKNHQRDPKRSSTFONLGPPLSLIHGTGSMQGLIGTVTVSNIVDIQT
VGLSTQEPGVFYAEFDGILGMAVPSLASEYSLPVDNMNHRLLVAQDLFSYMKRN
GOESMLTGLAINPSYTGSLHMYVTVQFQDVTQVLSVTSISGVVACEGSCQALIDTG
TSKIVGSSDILNIOALIGATONOGGFDIDCNLSYMPVVEINEMKPLPMSAVT
SDQGFCTSGFQSEHNSQKWLIDVIFIREYYSVFDNRNNLVGLAKAI"
29..773
/note="presequence"
200..1168
/product="chymosin b"
mat_peptide 305 a 393 c 340 g 267 t
BASE COUNT 305 a 393 c 340 g 267 t
ORIGIN
pst-1 site.
Query Match 76.8%; Score 900.4; DB 4; Length 1305;
Best Local Similarity 87.8%; Pred. No. 5e-263;
Matches 982; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 56 tcgtgtcgttactcaagctgctgagatcacgcacatctctctacaagaagtgaatctc 115
|||||
Db 54 TCTTGTGCTCTCTCCCAAGGCGCTGAGATCACAGAGATCCCTCTTACAAAGCAAGTCTC 113
QY 116 tccgtaagcgctgaagaaatgactctagaagaactcttcagaaacaacagatg 175
|||||
Db 114 TGAGGAAGCGCCTAAGAGAGATGGCTTCTGGAGAGACTTCTCAGAAACAGCATATG 173
QY 176 gcatcagcagcaagtaactccggtctcggtgaagttgtagcgttgccacttaaccaactacc 235
|||||
Db 174 GCATCAGCAGCAAGTACTCCGGCTTCCGGAGAGTGGCCAGCGTGCCTCCATCAACTAC 233
QY 236 ttgtagtcaataactttgggagaagatcttaacctcggaaccccgctccaagagtacagctc 295
|||||
Db 234 TGGATATGACTACTTGGGAAAGATTACCTCGGAGACCCGCCACAGAGATTACACCGTGC 293

Qy	296	tcgatactacggttccctcttactcttgggtctccctatctactgtgaagagcaatgct	355
Db	294	tgtttgacactggtccctctgacttcttggtgtaacctctatctactgcgaagcaatgct	353
Qy	356	gcaagaaacccaccaagatctcgatccgagaagtctgtccacctccagacttaagcaaac	415
Db	354	gcataaaacccacagcggttcgaccgagaaagatgtccactctccactccagacttgagcaac	413
Qy	416	ccctgtctatacactaagtgatcaggttagcatatgcagaagaaatttggctatgatactgtca	475
Db	414	ccctgtctatctacactacgagcagagcacatgcagagggcatctttaggctatgatacctgtca	473
Qy	476	ctgtctccaacatctgtgacatccaacagacagtagagcttagaaccccaagaacagatg	535
Db	474	ctgtctccaacatctgtgacatccaacagacagtagagcttagaaccccaagaacccgggg	533
Qy	536	atgtctcaacatctgcagaattcgatggcatctctgtgtatgtacataccatcgcgtcgt	595
Db	534	acgctcttcacctatgcccgaatttcgacgggagatccctgggagatggccttacccttcgcgcct	593
Qy	596	cagagtagctgcatacctctgttttgaacaacatgataagaaacagcaacttagtactaaagct	655
Db	594	cagatctactcatatccctctgtttgacacatgatagataagcagccacttggtggcccaagacc	653
Qy	656	tgtctcgttttaactatgcagaaatggtccagagagacatgtctcaagcttggaaactatg	715
Db	654	tgttctcgttttactatgcagaaatggtccagagagacatgtctcaccctggggggccatca	713
Qy	716	atccatctactatacagaatctcttcaacttcttcaactgttccagctgttgaagaaagtactg	775
Db	714	accctctctactatcacagaatctcttcaacttcttcaactgttgaagaaagtactg	773
Qy	776	aattcacgttgcagactgtccacatcaagcgtgtgtgttctgtcatgtataagtgatgtc	835
Db	774	agttcacgttgcagactgtccacatcaagcgtgtgtgttctgtcatgtataagtgatgtc	833
Qy	836	aagatacttggataccggtatcgcctcaagctgtgtctgcgaactagcagagcaattctcaaca	895
Db	834	aggccatcttggataccggtatcgcctcaagctgtgtctgcgaactagcagagcaattctcaaca	893
Qy	896	ttcagaagaactatgtgagacacacagaacacagtaacgtgtgatttgaatagtatgcaga	955
Db	894	tccacacagaccatttggagaccacacagaacacagtaacgtgtgatttgaatagtatgcaga	953
Qy	956	accttagcttaactgcctacagttgtcttcttgaatcaacggaagaatgtaccactgaccc	1015
Db	954	accttagcttactatgcctacagttgtcttcttgaatcaacggaagaatgtaccactgaccc	1013
Qy	1016	cctcgagctataccgagcagaatcaaaaggtcttgcacacagctggaatttcaaggttgaacc	1075
Db	1014	cctcgagcttataccgagcagaatcaaaaggtcttgcacacagctggaatttcaaggttgaacc	1073
Qy	1076	attcccaagaatgtatctcttggagagatgtgttcatctcgttgaatactactacgctcttga	1135
Db	1074	attcccaagaatgtatctcttggagagatgtgttcatctcgttgaatactactacgctcttga	1133
Qy	1136	ggggccaacaactcgttggcttagctgaagaactgtga	1173
Db	1134	ggggccaacaactcgttggcttagctgaagaactgtga	1171
RESULT 4			
LOCUS	AR002347	1240 bp	DNA
DEFINITION	Sequence 2 from patent US 5741665.	linear	PAT 04-DEC-1998
ACCESSION	AR002347		
VERSION	AR002347.1	GI:3963901	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1240)		
AUTHORS	Kato, E.K. and Stuart, W. Dorsey.		

[illegible]

	QY	956	accttagctacagtcgtcctcaagtgcgttcttgagaatacaaggaaatgtaaccacgagcc	1015
	Db	1000	ACCTGAGCTACTACTGCCCACTGTGTCITTTGAATCATATGGCAAATGTATCCACTGCACC	1059
	QY	1016	cctcgactataccagcagcagaatacaaggltctgtcacccagctgatccagatggaacc	1075
	Db	1060	CCTCGCGCTATACAGGCCAACAGACAGCGGCTTGTGTACCAAGTGCGCTTCCAGATGAAATC	1119
	QY	1076	attcccagaatatgatccttggagatggttcattcgttgagtaactaacagcgtctttgaca	1135
	Db	1120	ATTCCACAGAAAAGATCCCGGGGGATGTTTTATCCGAGAGTATTACAGCGCTTTTGACA	1179
	QY	1136	gggcacacacactcgttggtgctgactagctaagaacatctga	1173
	Db	1180	GGGCCAACACCTCGTGGGCTGGCCAAAGCCATCTGA	1217
RESULT	5	-		
LOCUS	E00042	1460 bp	RNA	linear PAT 29-SEP-1997
DEFINITION	E00042	DNA coding of pre-pro rennin.		
ACCESSION	E00042			
VERSION	E00042.1	GI:2168348		
KEYWORDS	JP 1982141287-A/1.			
SOURCE	unidentified.			
ORGANISM	unclassified.			
REFERENCE	1 (bases 1 to 1460)			
AUTHORS	Baanadetsule,R.A., Jienl,M., Donarudo,T.M., Arison,T.R. and Jierardo,F.B..			
TITLE	RENIN, PEPPERENIN OR PRORENIN GENE OBTAINED FROM RECOMBINED DNA MATERIAL AND LIVE CELL CONTAINING GENE Patent: JP 1982141287-A 1 01-SEP-1982;			
JOURNAL	KORABORATEIBU RES INC OS calf			
COMMENT	PN JP 1982141287-A/1 PD 01-SEP-1982 PF 14-JAN-1982 JP 1982003556 PR 16-JAN-1981 US 81 225717, 01-DEC-1981 US 81 325481 PI BAANADETSULE RABUTSUITSUKII ARUFUODOO, JIENI MAO, PI DONARUDO TETRAA MOJAA, PI ARISON TAUNTON RIGUBII, JIERARUDO FURANSHISU BUOBUSU PC C12N1/00,C07G67/00,C07H21/04,C12N15/00,C12P21/00//C12N1/18, PC C12N1/20, CC C12P19/34,C12R1/19,C12R1/865; CC strandedness: Double; CC topology: Linear; CC hypothetical: No; CC anti-sense: No; CC *source: tissue=stomach; FH key Location/Qualifiers FT CDS 205..1350 /product='pre-pro rennin' FT mat_peptide 205..1347 /product='pre-pro rennin' FT CDS 253..1350 /product='pro rennin' FT mat_peptide 253..1347 /product='pro rennin' FT CDS 278..1350 /product='rennin' FT FT mat_peptide 278..1347 /product='rennin'. FEATURES source Location/Qualifiers 1..1460 /organism='unidentified' /db_xref='taxon:32644'			
BASE COUNT	328 a	440 c	398 g	294 t
ORIGIN				

Query Match	76.5%;	Score 897.2;	DB 6;	Length 1460;
Best Local Similarity	87.7%;	Pred. No. 4.7e-262;		
Matches 980; Conservative	0;	Mismatches 138;	Indels 0;	Gaps 0;

OY	56	tcgttgctgttaaccacagcgtctggaataaccacgaatctctctcttcaaaaggaatcttc	115
Db	233	TCCTTGCGCTCTCCACAGGGCGCTAGATCACAGAGATCCCTCTGTACAAAGCAAGTCTC	292
OY	116	tcctgaagcgctgaaggaacatggaatctctagaagaactctctcagaacaacagtatg	175
Db	293	TGAGGAAGGCGCTGAAGAGACTGGGCTTGTGGAGGACTCTCGAGAAACACAGCTATG	352
OY	176	gcatacagaagaatactccgagctcgggtgaagtgtgtagctgagccacttaccactaac	235
Db	333	GCATCAGGAGAACTACTCCGGCTTCGGGAGGGGCGACCGTCCCTCCTACCACTATCC	412
OY	236	ttgatagtaacatactctgggaaga tctacactcggaaaccccgctcaagaattcacgcttc	295
Db	413	TGGANAGTCACTACTTTGGGAGAGTCTACTCGGGAGCCCGCCCGCCAGAGATTACCGTGC	472
OY	236	tccttgataactggttcctctgactctcgtgttccctctatctac tgcgaagaagaaatgct	355
Db	473	TGTTTGACACTGGGCTCCCTGTACTTCTGGGTACCCCTCTATCTCAAGAGCAATGGCT	532
OY	356	gcaagaaccaccaagaagtctgaatccggaagaatgcgtccactctcgaagaacttaggaaac	415
Db	533	GCAAAACCCACACCGCTTGACCCCGGAAGTGTCCACTTTCAGAACCTGGGCAAGC	592
OY	416	ccttgctatatacactagcgtgaatagtatgca tgaagaagaactcttaggctatgatccgtta	475
Db	593	CCCTGTATCTCACTACGGGAGACAGCAGATGACGAGCATCTGGGCTATGACACCGTCA	652
OY	476	ctgtctccacaactgtggaacatccaagacagtaagaacttagcaccccaagaacagtg	535
Db	653	CTGTCTCCAACTTGTGGACATCCAGCAGACAGTAGGCCCTGACACCCAGAGCCCGGG	712
OY	536	atgtcttcaacctatgacgaattctgaatgcaatccttgtagtgaataccatcgtcgtgt	595
Db	713	ACGTTCTTCACTATGCCGAATTTGCACGGGATCCTGGGGATGGCTTACCCCTGGCTGCCT	772
OY	596	cagagtaactcgataacctgtgtttgacacaca tga tgaaccgcacactagtaactcaagaact	655
Db	773	CAGAGTACTCGATACCCGTGTTTGACAACTGATGAACAGCAGCACGTGGGCCCAAGACC	832
OY	656	tgcttcctcggttataatggaacagaatgagccagaagaagatgcctccacagcttgaagctatg	715
Db	833	TGTTCTCGGTTTACATGAGACAGGAGATGGCCGAGAGACATGCTCACGCTGGGGCCATGG	892
OY	716	atccatccatacagaagatctcttcaactgggttccaagtaactgtgcagcagtaactgac	775
Db	893	ACCCGTCTTACTACACAGGGTCCCTGTGACTGGGTGCCCGTGACAGTGCACGACTACTGCC	952
OY	776	aatcaacbtggaacagtgatcaccaatcaacgagtggtgtgtgttgcatagttgaagtgatg	835
Db	953	AGTTCACTGTGAGACAGTGTACACATCAGCGCGTGTGGTTGGCCCTGTGAGGGGCGCTGC	1011
OY	836	aagatcatcttggaataccggaagtgatccaagtggttgagactagcaagcaatcttccaaca	895
Db	1013	AGGGCATCTCGGAGACGGGACCTCTCCAAAGTGTGGGGCCCGACACGACATCTCTAACA	1077
OY	896	ctcagcaagatcatctggagacacacagaacacagtaagtgagtttgacatagattgcgcga	955
Db	1073	TCCAGCAGGGCATTTGGAGCCACACAGAACAGTACGATGTGTGACATCGACTCGGACA	1133
OY	956	accttagctatacgtcctacaagtgctcttgagatacaagggcaagaatgataccaactgacc	1015
Db	1133	ACCTGACCTCATATCCCACTGTGGTCTTTGATGTCATATGGCAAAAGTCAACCACTGACCC	1192
OY	1016	ccttcgcgctataccagcagaagatccaaggtctctgcacacagtgatataccagagtggaag	1077
Db	1193	CTCTCGGCTTACACAGGCACACACAGGCGCTTCTGTACACAGTGGCTTCCAGATGAAATTC	1251
OY	1076	attcccaagaatgagatctctggagagatgtgttcaatctcgtagtaactacagcgtctttgaca	1135

|||||
Db 1253 ATGCCAAGAAATGATACCTGCGGGATGTTTTCATCCAGAGTATTAACGCGTCTTGACA 1312
Oy 1136 gggcacaacacccctggtgctagctaaagaatctga 1173
|||||
Db 1313 GGGCAACAACCTCGTGGGGCTGGCCAAAGCATCTGA 1350
|||||
RESULT 6
Al5836 Al5836 1291 bp DNA linear PAT 16-FEB-1994
LOCUS chymosin gene.
DEFINITION Al5836
ACCESSION Al5836
VERSION Al5836.1 GI:488959
KEYWORDS
SOURCE .
ORGANISM cow.
Bos taurus
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1291)
AUTHORS Simons,A.F.M. and De Vos,W.M.
TITLE DNA fragments, containing a lactic acid bacterium-specific
heterologous proteins
regulator region for the expression of genes coding for normally
Patent: EP 0307011-A 5 15-MAR-1989;
JOURNAL NEDERLANDS INSTITUUT VOOR ZUIVELANDERZOEK
FEATURES
source 1..1291
location/Qualifiers
BASE COUNT 323 a 382 c 328 g 257 t 1 others
ORIGIN
Query Match 76.4%; Score 896.2; DB 6; Length 1291;
Best Local Similarity 87.6%; Pred. No. 9.5e-262;
Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
Oy 56 tcgtgtcgttactacacgtgctagatcacccgcatctctctatacaagaagtaagtctc 115
|||||
Db 9 TCTTCGCTCTCTCCAGGGGCTGAGATCACCAGATCCCTCTGTACAAAGCAAGTCTC 68
|||||
Oy 116 tcgttaagcgtctgaaggaacatgactctagaagactctcttcagaacaacagatg 175
|||||
Db 69 TGAGGAAGCGCTGAGAGAGATGGGCTTCTGAGAGACTTCTCGAAGAAAGCAGTATG 128
|||||
Oy 176 gcatcagcagcaagtaactcggcttggtgaagtgtcgaagtgtcgaacttaccatc 235
|||||
Db 129 GCATCAGCAGCAAGTACTCCGGCTTCGGGAGGTGGCCAGCGTGCCTTGACCAACTAC 188
|||||
Oy 236 ttgataagcaatcttggaagaatctaccccgaaacccctcaagaagtcaacgttc 295
|||||
Db 189 TGGATGATGACTACTTGGAGATCTACTCTGGGACCCGCCCAAGAGTTCAACCGTGC 248
|||||
Oy 296 tcttgataactggtctcctctactctcttggttcctctatactactcaagaagcaagtc 355
|||||
Db 249 TGTGTGACACTGGCTCTGACTTCTGGGTACCTCTATCTACTGCAAGACATGCTCT 308
|||||
Oy 356 gcaagaaccacaagaattcgatccgagaagtcgtcaaccttcagaacttaagcaaac 415
|||||
Db 309 GCAAAAACCCAGCGCTTCGACCCGAGAAAGTCTCCACCTTCAGAACCTGGGCAAGC 368
|||||
Oy 416 ccttgtataactaagtagataggttagcatgaagaacttaggctaataaccgtca 475
|||||
Db 369 CCTGTGTATTCATACGAGACAGCAGCATTCAGGGCATCTGGGCTATGACACCGTCA 428
|||||
Oy 476 ctgtctcaacatctgtagacatctcaacagagtagaacttagcaccacaagaacagagtg 535
|||||
Db 429 CTGTCTCAACATTTGTGGACATCCAGACAGTAGGCTTGACACCCGAGGCCGGGG 488
|||||
Oy 536 atgtcttaacctatgcaagaattcgatgcatccttggtatgtagcaaccaatgcgtcgt 595
|||||

Db 489 ACGTCTTACCTATGCCGAATTGACGGGATCCTGGGGATGGCTACCCCTCGCGCT 548
Oy 596 cagagtaactcgaatacctglttgtaacaacatga tgaacgcacacctaagtaagtaagta 655
|||||
Db 549 CAGAGTACTCGATACCCGTTGTGACACATGATGAACAGGACACCTGTTGGCCAAAGACC 608
|||||
Oy 656 tgtctcgttctacaatgaagaatgagccagaagatgcatgctcgaagcttgagagctatg 715
|||||
Db 609 TGTTCGTGGTTTACATGAGACAGGAATGGCCAGAGACATGCTACGCTGGGGCCATCG 668
|||||
Oy 716 atccatcctactacaagaatcctcctcaactggttccagactgtagcagaatgaatgagc 775
|||||
Db 669 ACCGNTCTTACTACACAGGGTCCCTGACTGCGGTGCCCGGACAGTCCACAGTACTGCGC 728
|||||
Oy 776 aattcactgtgacagtgatccatcagcagtggtgtgtgtgtgtgtgtgtgtgtgtgtgt 835
|||||
Db 729 AGTTCATGTTGACAGTGTACAGATCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 788
|||||
Oy 836 aagctatcttgataacccgtagctcaagcttggttgcgaacttagcagagatctcaaca 895
|||||
Db 789 AGGCAATCTTGACACCGGACACCTCCAAAGCTGTGCGGGCCAGACGACATCTCTACACA 848
|||||
Oy 896 ttcaagaagctatgtagacacacacagacacagataggttgatgtatagatagatgtagaca 955
|||||
Db 849 TCCAGCAGGCCATTTGAGACCAACAGACAGACAGTACGGGTGATTTGACATGACAGTGCACA 908
|||||
Oy 956 accttaagctatagctcctacagtgltcttgagatcaacagcagaatgtaccactgacc 1015
|||||
Db 909 ACTGAGCATACATGCCACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 968
|||||
Oy 1016 ccttcgctatatacagcagacagatcaaggtgtctcaccagtgtatccagatgaagacc 1075
|||||
Db 969 CCTCCGCTATACCAAGCAGGACCAAGGCTCTCTTACCAAGTGTGCTTCCAGAGTAAATATC 1028
|||||
Oy 1076 attccagaatgagatcttgtagatgtgtcattctgtagaactacagcgtcttgaca 1135
|||||
Db 1029 ATTCGCAAGAAATGATCTCGTGGGATGTTTCATCCGAGATATTACAGCGCTTTTGACA 1088
|||||
Oy 1136 gggccaacaacctcgttggtgcttagctaaagaatctga 1173
|||||
Db 1089 GGGCCAAACACTGTGTGGGCTGGCCAAAGCCATCTGA 1126
|||||
RESULT 7
I04058 I04058 1175 bp linear PAT 02-DEC-1994
LOCUS Sequence 5 from Patent EP 0123928.
DEFINITION I04058
ACCESSION I04058
VERSION I04058.1 GI:591912
KEYWORDS
SOURCE .
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1175)
AUTHORS Cashion,L.M., McCaman,M.T., Rice,C.W. and Sias,S.R.
TITLE Recombinant DNA coding for a polypeptide displaying milk clotting
activity
Patent: EP 0123928-A2 5 07-NOV-1984;
JOURNAL Location/Qualifiers
FEATURES
source 1..1175
location/Qualifiers
BASE COUNT 263 a 356 c 318 g 238 t
ORIGIN
Query Match 76.4%; Score 895.6; DB 6; Length 1175;
Best Local Similarity 87.6%; Pred. No. 1.4e-261;
Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
Oy 56 tcgtgtcgttactacacgtgctgagatcacccgcatctctctatacaagaagtaagtctc 115
|||||
Db 49 TCTTCGCTCTCTCCAGGGGCTGAGATCACCAGATCCCTCTGTATCAAAAGCAAGTCTC 108
|||||

OY	116	tcctgataagccgtttgaaggaacaatgagactcttcaagaagatctcttgcacaaacaacagatg	175
Db	109	tcagcaagagccgttgaggagacatggcgcttctgagagacattcttcacaaacacagatgatg	168
OY	176	gcataagcagcaagfacticcggtctcggttgaagtctgacgctgtccacttaacaactacc	235
Db	169	gcattcagcagcagatctccggcttccggggagagtgccagcgctcccttaccacactacc	228
OY	236	ttgatatgcataactttggaagaltactctcggaaccccgctcaagaltcaacgctc	295
Db	229	tggatagtcagacttcttgggaagatctaccctcgggaccgcccccacagattcacctgc	288
OY	296	tctttgatacttgatccctctgactctctggtctccctatratctatgtgaagaagtgct	355
Db	289	tgttttgacacttgctcctctctgactcttctggataccctctattatgtgaagaagattgctt	348
OY	356	gcaagaacaccaccaagaaltcgatccgagaagtgcttcacccttcagaacttaggcaaac	415
Db	349	gcataaaccaccagcggttcgaccocgagaagtgctccaccttcacaaactggcgaaac	408
OY	416	ccctgtctataactacagtgatcaggttagcatgtcaaggaatcttagctatgataccgtta	475
Db	409	ccctgtctatctccacacacgagacagacacacgagccatcttggttatgacacactgca	468
OY	476	ctgtctccaaacttgtagacaaltcaaacagaagataggacttagaacccaagaacaggtg	535
Db	469	ctgtctcccaaacatttgtagacattcaaaagagacatgagaccttgatgacacccgggg	528
OY	536	atgtcttcaactatgacgaaltctgatalgacalccttggtatgtgacataccaatgctcgct	595
Db	529	acgtttccactatgaccgaatttcacacgggattcctggggatggacctacccttcgctgcct	588
OY	596	cagagtatctgatactcgttgtttgacaacatgatgaaccacacactgtagtccaaagct	655
Db	589	cagactactcgaattccctggttttaacaacaatgatgaacacagacactgtagtccaaagac	648
OY	656	tgtctcgttttatactgacacaggaatggccagagagacatgatacagcttgtagagctattg	715
Db	649	tgttctgcgttttaactgtagacagaaatggccagagagacatgctacgcttggggccatgc	708
OY	716	atccatctctaactacaagatctcttcaactggtgtccagtaactgtgacagtaactgagc	775
Db	709	accctctctactacacagaggtcccttgacactggggcccgtagacatgtaacaaactgtgagc	768
OY	776	aattcaactgtgacagtgatcaactcaagcgtgtgtgtctgtcatgtgaagtgatgagatgc	835
Db	769	agttcactgtgacacagtgtaacacatfcaacgggtgtgtgtgtgtgtgtgtgtgtgtgtgt	828
OY	836	aagatacttggataccggtatcgttccaaagctgtgtcgaacctagacgaacgaattctcaaa	895
Db	829	aggccatctcgtgagacacggggcactcccaagctggttcggggccacagacgaatctcaaca	888
OY	896	ttcagcaagctattggtgggcccacaagaacacagtaacggttgatgttgacatagatitcgaca	955
Db	889	tccacacagccatttgagaccacacagaaacagtaacgattgttgacatgcattggagaca	948
OY	956	accttagctatactccatacaattgctcttgatagatacgaagcagaatgtataccactgacc	1011
Db	949	accttagctatactccatacaattgctcttgatagatacgaagcagaatgtataccactgacc	1001
OY	1016	ccctcgactataccagccaggaatcaaggttctgtcacacagttgattccagatgtgaacc	1077
Db	1009	cctcgactataccagccaggaatcaaggttctgtcacacagttgattccagatgtgaacc	1061
OY	1076	attcccaagaatgtatctcttggaatgtgttcaatctgtgtgatacttaacaggtcttttgaca	1133
Db	1069	attcccaagaatgtatctcttggaatgtgttcaatctgtgtgatacttaacaggtcttttgaca	1122
OY	1136	gggccaacaacactcgtttggtcgtatgacttaagaacatctga	1173
Db	1129	gggccaacaacactcgtttggtcgtatgacttaagaacatctga	1166

RESULT	8	AR073077	2733 bp	DNA	linear	PAT 28-AUG-2000
LOCUS		AR073077				
DEFINITION		Sequence 6 from patent US 5948682.				
ACCESSION		AR073077				
VERSION		AR073077.1	GI:9999840			
KEYWORDS						
SOURCE		Unknown.				
ORGANISM		Unknown.				
REFERENCE		1 (bases 1 to 2733)				
AUTHORS		Moloney, M. M.				
TITLE		Preparation of heterologous proteins on oil bodies				
JOURNAL		Patent: US 5948682-A 6 07-SEP-1999;				
FEATURES		Location/Qualifiers				
	source	1..2733				
		/organism="unknown"				
BASE COUNT		699 a 681 c 684 g 669 t				
ORIGIN						
Query Match		76.3%; Score 894.8; DB 6;				
Best Local Similarity		88.4%; Pred. No. 2.7e-261;				
Matches	971; Conservative	0; Mismatches 127; Indels	0; Gaps	0;		
QY	76	gctgagatcacccgcattccctctctctcaaaagtgtaagttcttcgcgttaagcgctgaagaa	135			
Db	1630	GCATGAGATCACACGAGATCCCTCTGTACAAAGCAAGTCTTGAGAAAGCCCTGAGAG	1689			
QY	136	catgactcttcaagaagactctcttcagagaacaagaatgcatcagagaagaatattcc	195			
Db	1690	CATGGGCTTCTTGAGAGATTCTCTGCAAGAAAGCAGTATGCGATCAGCAGCAAGTACTCC	1749			
QY	136	ggcttcggltgaagttgctagcgtgtgcacttaccactactcctttagtactattctggg	255			
Db	1750	GGCTTGGGGAGGTGGCCAGCGTGCCCTGTGACCAACTGATGATGATGATGATGGG	1809			
QY	256	aagactcacccttcggaaccccgctcctaagaagttcaacggttctctttgatacagttccctc	315			
Db	1810	AAGATCTACCTCGGAGACCCCGCCCGCAGAGATTCACCCGTGTGTGACACAGGCTCTCT	1869			
QY	316	gaactcttggttccctctcatctactctcaagaagcaatgctctcaagaacaccacaagaatc	375			
Db	1870	GACTTCTGGGTATCCCTGTATCTAGTGCAGAAAGCAATGCTGTGCAGAAAACACACGCGTTC	1929			
QY	376	gaccgagaagaagtgcgtccaccttcagaacttaggcaaaaccctgtctatacactacggtc	435			
Db	1930	GACCCGAGAAAGTCTCTCACCTCCAGAACCTCGGGGACGCCCTGTGTATCTACCTACGGG	1989			
QY	436	acaagttagcatgaagaagatcttaggctatgataccgttactgtctccaaatctgtgac	495			
Db	1990	ACAGGCACACATGCGAGGCAATCTTGCGCTATATACACCGTCACTGTTCTCCAAATGTGGAC	2049			
QY	436	attcaacagacaagtagaacttagcaccaccaaaacacaggtgtagtctctcaactatgcagaa	555			
Db	2050	ATCCAGCAGACAGTATGGGCTTGAGCACCCAGAGAGCCGGGGAGCGTCTACACTATGCCGA	2109			
QY	556	ttcgatgcatcccttgyratatggtacatacccatcgctcgtcgtctaaagtagtactcgatc	615			
Db	2110	TTTCAGCGGAGATCTGGGGATGGCTTACCCCTCGCTCGCTTAGAGTACTGATATCCCGTG	2169			
QY	616	tttgacaacatgataaaccgaacactagtagtctaaagaactgttctcggtttataatgac	675			
Db	2170	TTTGTACACATGATGAACAGGCACTGTGTGCCCCAAGAACCTGTTCTGCGTTTACATGGAC	2229			
QY	676	aggaatggccaagagacatgctcaagcttggagatattgatacactactacacagaa	735			
Db	2230	AGGATGGCCAGGAGAGATGCTCACGCTGGGGGCCATATCGACCCGCTCTACTACACAGGG	2289			
QY	736	tctcttaacttggttctcaagctatgctatgctgaagaatgacgaatctatcttggaaaggtc	795			
Db	2290	TTCCTGCACTGGGTGGCCGATGACAGTGAAGGATCTGGGCACTTCTGAGACATGATC	2349			

Oy	796	acacaaagcagtgatggttgatgcacatgaaagtgtgaatgcaagctatcttgataccgt	855
Db	2350	ACCATCACCGGCTGTCGTGTGTGACCTGTGAGGTGGCTTCAGGCCATCTTGGACACGGGC	2409
Oy	856	acgtccaagctgcttgagaccctagcagcagacattctccaacttcagcaagctattggaacc	915
Db	2410	ACCTCAAGCTGTGTGGGGCCGACAGCGACACATTCCTCAACATTCGACAGCCATTGGACC	2469
Oy	916	acacagaaccagctacgctgtagcttgaacatagattgagacaacctagctacatgctctaca	975
Db	2470	ACACAGAACCCAGTACGGTAGCTTTGACATTCACATCGACACCTGAGCTACATGCCCCACT	2529
Oy	976	gttgctcttgagatcaaggaagaatgacccactgacccctccgcttatacaagccag	1035
Db	2530	GTGGCTTTGAGATCAATAGCGCAAAATGTATCCCTACGACCCCTCCGCTATATACAGCCAA	2589
Oy	1036	gatacaaggtctgcaccagctgattctccagatgtaaacattccacgaatgtgactctg	1095
Db	2530	GACCAGGGCTTCTGTATCCAGTGGCTTCCAGAGTGAATAATCATTTCCAGAAATGATCTCTG	2649
Oy	1096	ggagatgtgttcatctcgtgagctactacacgctcttggacagggccacaacctgcttgg	1155
Db	2650	GGGAGATGTTTTCATCTCGAGAGATATTACAGCGCTTTGACAGGGCCAAACACTCTGTGGGG	2709
Oy	1156	ctagctaaagcaatcttga 1173	
Db	2710	CTGGCCAAAGCCATCTGA 2727	

LOCUS	BOVCHWMOA	1275 bp	mRNA	linear	MAM 26-APR-1993
DEFINITION	bovine chymosin a (rennin) mRNA.				
ACCESSION	J00002				
VERSION	J00002.1	GI:162857			
KEYWORDS	chymosin; chymosin A; rennin.				
SOURCE	bovine (calf) cdna of fourth stomach mucosa mRNA.				
ORGANISM	Bos taurus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.				
AUTHORS	Molr,D., Mao,J.I., Schumm,J.W., Vovis,G.F., Alford,B.L. and Tauton-Rigby,A.				
TITLE	molecular cloning and characterization of double-stranded cna coding for bovine chymosin				
JOURNAL	Gene 19, 127-138 (1982)				
MEDLINE	83054629				
COMMENT	chymosin (rennin) is the major proteolytic enzyme in the fourth stomach of the unweaned calf. two chromatographically distinct forms are known and a third seems likely (see bovchymob and bovchymoc). this sequence has been tentatively identified as chymosin a; it differs from chymosin b by only two amino acids and from chymosin c by another amino acid. [1] argues that the different chymosins are probably polymorphic variants of a single chymosin gene.				
FEATURES	Location/Qualifiers				
source	1..1275				
	/organism="Bos taurus"				
	/db_xref="taxon:9913"				
	21..1166				
	/note="preprochymosin a"				
	/codon_start=1				
	/protein_id="AAA30447.1"				
	/db_xref="GI:162858"				
	/translation="MRCVLAVLAVPALSOGAEETRIPIYKCKSLRKALKEHQLLEDFT				
	OKOQYISKYSGEFPEVASVPLTNTLNDISQYFQK IYLGPPQFVLPFTGSSDFWVP				
	IYCKSNACKNHORFPRKSPFQNTGKSLSHYSGMGLGYDTVYSNLIDLOQ				
	VGLSTQPEGDFTVAEPGIIIGMAYPLASEYISIPFNMMNRHLVADDFSVYMDR				
	GOESMTLIGATIDPCYTGSHHWNPVTVQOQMFQVDSVTSISVVVACGGCOALIDDT				
	TSKLVPSSDILNTIOAIGATGNQDDEDDICDNLSTPYTFEINCKMPLPLPSATY				
	SDDQCFSTSGFQSESHNSQWMLGDVFIREYTSVPRANNVLGLAKAI"				
	24..68				

mat_peptide	/note="presence"
	195.1163
BASE COUNT	/product="chymosin a"
	293 a 391 c 336 g 255 t
ORIGIN	20 bases upstream from codon 1.

Query Match	76.2%	Score 894	DB 4	Length 1275
Best Local Similarity	87.5%	Pred. No. 4.4e-261		
Matches 978	Conservative	0	Mismatches 140	Indels 0
Oy	56	tcgtgtcgtgtctctcaacgctctcgtatgataccgcatctccctctcaaaaggttaagcttc	115	
Db	49	tctttgcctctctccacggcgcttgatgataccagagatccctctgtatcaaaagcgaatctc	108	
Oy	116	tcgcgtaaagcgtctgaagaaacatgtgactctctgaagaactctcttgcaagaacaacgttgc	175	
Db	109	tcgaggaagcgcgttgaaaggagcattggcgcttctggagagacttctctgcagaagcagctatg	168	
Oy	176	gcatgaagcagaagtaactccgctctcgtgtgaagttgtacgctgtgcaacttacaactacc	235	
Db	169	gcattcagcagcagaatctccgctctcggggagagtgccacgcctgccccttgacacactacc	228	
Oy	236	ttgatagtcaactcttgggaagatctaccctcggaaeccccgcctcaagaagttcacgcttc	295	
Db	229	tggatgagtcagacttttgggaagatctactctcgggaccccgcctcagagattccagcttc	288	
Oy	296	tctttgatactgttctctcgtactctcgttgcctctctctctctctctctctctctctctct	355	
Db	289	tgtttgacactgctcctctctgactcttggggatccctcttcttactgccaagacatgctt	348	
Oy	356	gcaagaaccaccaaagaatctcgaatcggaaagatgcgtccactccgaactgaagcaaac	415	
Db	349	gcamaaacaccagcgcttgcacccgagaaagtcgcccacttccagaaactcgggcaaac	408	
Oy	416	ccttgcgtatacactcgcgttacaagtgatgataagaaatcttagctctgtatcacgtca	475	
Db	409	ccctgtcttctcactcgttgcagagacagcagatgcagagggcactctcggctatgacaccgtca	468	
Oy	476	ctgtcccaacatctgtggacatctcaacagacagtaagactctagcacccaagaacacagtg	535	
Db	469	ctgtctccaaacattcttcgacatctcagatccacagacagatagccctgacaccacccgggg	528	
Oy	536	atgtctcaacctatgcagaatctcgatgtgcacctctgttatgycataccatcgctcgct	595	
Db	529	acgtcttccacgtatgccgaattgcaggggactcctggggatgggcttaccctcctcgctt	588	
Oy	596	cagaagtaactcgaactctgtcttttgcacaatgaatgaacgacacactagtaactcacaagt	655	
Db	589	cagaagtaactcgaatcccgctgtttgacaaactatgtgaacagacacactgtggcccaagac	648	
Oy	656	tgctctcgtttaaactgaagaaatgycacaggaagacatgctcaacgctctgaagctatg	715	
Db	649	tgtttctcgtttacatgagacagaaatccacagagacagatgctcagcctggggccatcg	708	
Oy	716	atccatcctactacacaagaatctcttcacgtgggtcttcagttcactctgtgcaacagtaactgc	775	
Db	709	accctgctactacacacaggggtccctgcactgggtgcccctgacaaagtgcacgaatctgccc	768	
Oy	776	aattcactgtggacagtgtaaccatcaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	835	
Db	769	agtttcaactgtggacagtgtaaccatcaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	828	
Oy	836	aagctactcttgataccggttgcgttccaagctgtgtcggagcttaagcagcgaatcttcaaca	895	
Db	829	agcctacactcttgataccggttgcgttccaagctgtgtcggagcttaagcagcgaatcttcaaca	888	
Oy	896	tttagcaagatcttgtagcacacacagaacagcagaagttgagtttgatacatagattcgcaca	955	
Db	889	tccacagcagcctttagtagcacacacagaacacagatgagtttgatacatagactcgcaca	948	
Oy	956	accttagctacatgctctacagttgtctttttagatcaacggcgaagatgtaccactgaacc	1015	

Db 949 ACCTGAGCTACATGCCACTGTGCTTTGAGATCAATGCGAANAATGTACCCACTGACCC 1008
QY 1016 cctccgacctataccagccagatcaagggtctcgcaccagctgagatctcagaagatgagacc 1075
|||||
Db 1009 CCTCGGCTATACCAAGCCAGACGAGGCTCTGTACACAGTGGCTTCCAGAGTGAATAATC 1068
QY 1076 attccagaatgagatcttggagagatgcttcacatcgtgagtaactacagacgctcttgaca 1135
|||||
Db 1069 ATTCCAGAAATGATCTCTGGGGGATGTTTTCATCCGAGATATACAGCGCTTTGACA 1128
QY 1136 gggccaaacctcgttggctagctaaagcaatctga 1173
|||||
Db 1129 GGGCCAACACCTCGTGGGCTGGCCAAAGCCATCTGA 1166

RESULT 10
E00075
LOCUS E00075 1290 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding chymosin (=rennin).
ACCESSION E00075
VERSION E00075.1 GI:2168379
KEYWORDS JP 1983009687-A/1.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 1290)
Nomenclature: Walker, T.D., Chimoshi, J.R.H., Pilla, A.R. and
Nijon, S.E.
TITLE PRODUCTION OF POLYPEPTIDE
JOURNAL Patent: JP 1983009687-A.1 20-JAN-1983;
COMMENT CELTECH LTD
OS bovine
PN JP 1983009687-A/1
PD 20-JAN-1983
PF 17-JUN-1982 JP 1982104672
PR 17-JUN-1981 GB 81 8118688, 11-NOV-1981 GB 81 8133998, PR
01-DEC-1981 GB 81 8136185, 10-FEB-1982 GB 82 8203907 PI
NOOMAN HENRI KEARI, MAIKERU TEREUSU DOORU,
PI CHIMOSHI JYON ROI HARISU
PI PIITA ANSONII ROU, JYON SUPENSAA EMUTEEI
PC C12M9/52, C07H21/04, C12N1/00, C12N9/60, C12M15/00, PC
C12Q1/02, C12R1/19,
PC C12R1/865;
CC strandedness: Double;
CC topology: linear;
CC hypothetical: No;
CC *source: tissue: stomach;
CC *source: clone: PCT 70;
FH Key Location/Qualifiers
FH 5'UTR 1..25
FT sig_peptide 26..73
FT mat_peptide 74..1078
FT 3'UTR /product='pro-chymosin'
FT CDS 1082..1200
FT 26..1081
FT /product='pre-pro-chymosin'.
FEATURES
Source Location/Qualifiers
1..1290
/organism='Bos taurus'
/db_xref='taxon:9913'
BASE COUNT 309 a 384 c 334 g 263 t
ORIGIN

Query Match 76.2%: Score 894; DB 6; Length 1290;
Best Local Similarity 87.5%: Pred. No. 4,4e-261;
Matches 978; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 56 tcgtgtgttactcaagctcgtgagatcacccgcatctctctacaagtaagtc 115
|||||

Db 54 TCTTCGCTCTCTCCCAAGGCGCTGAGATCACAGAGATCCCTCTGTACAAAGCAAGTCTC 113
QY 116 tccgtaagcgctgaggaatgagatctctagaagatctcttgagaagaacagtg 175
|||||
Db 114 TGAGGAAGCGCTGAGAGAGCATGGGCTTCTGAGAGACTTCCGTGAGAAACGCGATG 173
QY 176 gcatagagcaagtaaccggtctcgttgaagttgacgtgtgacatctcaactacc 235
|||||
Db 174 GCATAGAGCAGCAGTACCTCCGCTTGGGGAGGTGGCCAGCGTGGCCCTGACCAACTAC 233
QY 236 ttgataagtaactcttggagaagatctaccctcggaacccgcctcaagatctaccgltc 295
|||||
Db 234 TAGATAGTCACTATTGGGAAGATCTACCTGGGAGCCGCCCCAGAGATGTACACCTGTC 293
QY 296 tcttgatacgtgttccctcgtactctgtgtccctcctatctactgtaaggaatgct 355
|||||
Db 294 TGTGTGACACTGGCTCTCTGACTTGTGGATCCCTATCTACTGCAAGCAATGCTT 353
QY 356 gaaagaaccacaaagatctgataccgagaagctgtccacctccagaacttaagcaaac 415
|||||
Db 354 GCAAAAACACACAGCGCTTCGACCCGAGAAAGTGTCCACCTTCCAGAACTTGGGCAAGC 413
QY 416 cctgtctatacactaagtgatcaggtagatgaaagaatcttaggtatgatacgtca 475
|||||
Db 414 CCTGTCAATCCACTACCGAGCAGCAGCAGCAGGCGCATCTAGATATGACACCCCTTA 473
QY 476 ctgtctcaaatgttgaaatctcaacgagatgagacttagaccaccaagaacagtg 535
|||||
Db 474 CTGTCTCAACATGTTGAGATCTCAGCAGACAGTAGAGCTTAGACACCCAAAGCCCGGG 533
QY 536 atgtctcaactatgaaagatctgataccctgtgtatgataccatcgtcgtcgt 595
|||||
Db 534 ACGTCTTACCTATGCGCAATTCAGCGGATCTGTGGGATGGCTTACCTCGCTGCT 593
QY 596 cagagctacgatacctgtgttgacaacatgataagcaacacacttagactcaagact 655
|||||
Db 594 CAGAGTACTCATACCCCTGTTTGACAAATGATGAAAGCAGCAGTGTGGCCCAAGAC 653
QY 656 tgtctcgttctacatgagacaggaatggcagagagagatctcaacgttggagctatg 715
|||||
Db 654 TGTTCCTCGTTTACATGAGACAGATTTGTCAGGAGAGATCTCAGCTGGGGGCGATCA 713
QY 716 atccatctactaacaagaatctctcactgtgtcagctacgtgtgagcagactagtc 775
|||||
Db 714 ACCGCTCTACTACACAGAGTCCCTGACCTGGTCCCGGTGACAGTGCAGTACTGCG 773
QY 776 aatcactgtgagacagtgatccatcagcgtgtgtgtgtgtgacatgtgaagtgatgc 835
|||||
Db 774 AGTTCACGTGTGAGACAGTGCACATCACGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 833
QY 836 aagatctcttgatacagcgtatctccaaagctgtgtcgaacctagaagagacatctcaaca 895
|||||
Db 834 AGGCCATTTGGACACGGGACCTTCACAGCTGTGTGGGCCGACGACGACATCTCAACA 893
QY 896 ttccagaagctatgagacacacagaaacagtaacgtgtgagtttgaacatagatgagaca 955
|||||
Db 894 TCCAGCAGCATTTGGAGCCACACAGAACGATGCGGTGATTTGACATGACATGCGGACA 953
QY 956 acctagctacatcctcaagtgctcttgatgaatcaacgcaagatgtaccactgacc 1015
|||||
Db 954 ACCTGAGCTACATGCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1013
QY 1016 cctccgacctataccagccagatcaagggtctcgcaccagctgagatctcagaagatgagacc 1075
|||||
Db 1014 CCTCGGCTATACCAAGCCAGACGAGGCTCTGTACACAGTGGCTTCCAGAGTGAATAATC 1073
QY 1076 attccagaatgagatcttggagagatgcttcacatcgtgagtaactacagacgctcttgaca 1135
|||||
Db 1074 ATTCCAGAAATGATCTCTGGGGGATGTTTTCATCCGAGATATATACAGCGCTTTGACA 1133
QY 1136 gggccaaacctcgttggctagctaaagcaatctga 1173
|||||
Db 1134 GGGCCAACACCTCGTGGGCTGGCCAAAGCCATCTGA 1171

RESULT 11
LOCUS A15633 1291 bp mRNA linear PAT 18-FEB-1994
DEFINITION preprochymosin.
ACCESSION A15633
VERSION A15633.1 GI:491951
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 1291)
AUTHORS Carey,N.H., Doel,M.T., Harris,T.J.R., Lowe,P.A. and Emlage,J.S.
TITLE A process for the production of a polypeptide
JOURNAL Patient: EP 0068691-A 29 05-JAN-1983;
CELLTECH LIMITED
FEATURES
source Location/Qualifiers
1. 1291
/organism="synthetic construct"
/db_xref="taxon:32630"
26. 1171
/gene="preprochymosin"
26. 1171
/gene="preprochymosin"
/codon_start=1
/transl_table=11
/protein_id="CAA01244.1"
/db_xref="GI:491952"
/translation="MRCLVYLAVFALSGAETTRIPLYKSKLRKLRKHGLDEFL
OKQOYISKYSSEGFVAVPLTNVLDSDTFKIVGTAPQEFTEVLDDSSPWFPS
YCKSNACKNHFQFDRKSTFQNLGKPLSIHGTGSMOILIDYDVYNSIVDIOOT
VGLSTQEPGVDFYAEFDGILGMAYPSLASEYSIPEDNMNHLVAODLFSVMRI
GOSMTLGIADINPSTYGLHWVPYVOQWFOFTVDSVTISGVVAECGGCAILDG
TSLVGPSSDILNIOAIGATONOGFEDIDCDNLSPMPVVEINGKMPPLPSAVT
SDQGFCTSGFSEHNSOKWILGDVFIREFYVFPDANNVLGLAKAI"
misc_difference 1291
BASE COUNT 309 a 384 c 334 g 263 t 1 others
ORIGIN
Query Match 76.2%; Score 894; DB 6; Length 1291;
Best Local Similarity 87.5%; Pred. No. 4.4e-261;
Matches 978; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 56 tcattgctgttaccagctctagatcaccgccattccctcttacaagaagtatgctc 115
Db 54 TCTTGGCTCTCCCAAGGCGCTGAGATCACCAGATCCCTCTGTACAAAGCAAGTCTC 113
QY 116 tccgtaaggcgtgaaggaaatggaactctagaagaactcttgcaagaacaagatag 175
Db 114 TGAGGAAGGCGCTGAGAGAGATGGGCTTGTGAGAGACTTCTCGAAGAACAGCAGTATG 173
QY 176 gcatcagaagaagtactcggctctggttgaagtgtcagtggtgcactaccactacc 235
Db 174 GCATCAGCAGCAAGTACTCCGGCTTGGGAGAGTGGCCAGGCGCTCCGTGACCAACTATCC 233
QY 236 ttgatagtcaactcttggagaagatctaccccggaaccccgcccaagaatcagctgc 295
Db 234 TAGATAGTCACTACTTTGGGAAGATCTACTCGGAGACCCGCCCAAGAGATTACACCGTGC 293
QY 296 tctttgatactggtctcctctgactcttgggtctccctctatctactgcaagaatgctc 355
Db 294 TGTGTGACACTGGGCTCCTGACTTCTGGTACCTCTATCTACTGCAAGCAATGCTCT 353
QY 356 gcaagaaccaccaagaatgcatcagagaagaatgctcactctcagaacttaggcaaac 415
Db 354 GCAAAAACACACAGCGCTTGCACCCGAGAAAGTGTCCACCTCCAGAACTTGGGCAAGC 413
QY 416 cctgtctatcactaagtgtagttagcatgcaagaagaactttagctatatactacgtca 475
Db 414 CCCTGTCAATCCACTACGGACAGGACAGATGCAAGGCAATCCCTAGGATATGACACCGTTA 473

QY 476 ctgtctcaaacatctgtggacattcaacagacagtaggaacttagcacccagaagaagtg 535
Db 474 CTGTCTCCACAAATTTGTGACATCCAGACAGATAGGCTGTAGACACCCAGAGCCCGGG 533
QY 536 atgtcttcacctatgcagaatctgtagtctctcttgtaggaataccatctgctcgt 595
Db 534 ACGTCTTACCTTATGCGGAATTCGACGGATCTGGGGAGTGGCTTACCTCGCTGCTCT 593
QY 596 cagaagctacgatacctgtgttgtagacacatgataagacacacactgtagtccaagaact 655
Db 594 CAGAGTACTGATRCCGTGTGTTGACAAATGATGAAACAGGCACTTGTTGGCCAAAGCC 653
QY 656 tgtctcgttatactatggaagaatggaagagagagatgctcaogcttggagatctg 715
Db 654 TGTCTCGGTTTACATGAGACAGATGTGTCAGAGAGACATGCTCACCTCGGGGCAATCA 713
QY 716 atccatcctactaacaagatctcttacttgggttccagtcactgtgcaagaactgagc 775
Db 714 ACCCGTCTTACTACACAGGGTCCCTGCACTGGGTGCCGTGACGTGACAGTACTGCT 773
QY 776 aattcactgtggagacgtgcaaccatcagcgtgtgtgtgtgtcatgtgaagtgatgctc 835
Db 774 AGTTCACTGTGACAGTGTACACATCAGCGGTGTGTGATGTGAGAGGTGCTGCTC 833
QY 836 aagctatcttggataccggtacgtlccaagctgtgcgaacctagcagcaacttccaaca 895
Db 834 AGGCATCTTGGACACGGGGCACTCCAAAGCTGTGGGGCCAGCAGGACATCTCCACAA 893
QY 896 ttgacgaagctatggagacacacagagaacagatgacgttgaattggaatagatggaca 955
Db 894 TCCAGCAGGCCATTGGAGCCACACAGAACAGTACGGGTGATTTGACATGACACTGCGACA 953
QY 956 accttaagctacatgcctacagtgcttcttgaatcaacggaagaatgtaaccactgaacc 1015
Db 954 ACCGAGGTATATCCCACTGTGCTTGTGATGATCAATGGCAAAATGATACCCTGACCC 1013
QY 1016 cctcgcctataccagccagaatcaagggtctcagaccatgtaattccaagatgaagacc 1075
Db 1014 CCTCCGCTATACCAAGCCAAAGCAAGGCTCTGTACCAAGTGTCCAGATGTAATAATC 1073
QY 1076 attccagaagaatgagatcttggagaatgtgtcattccttgtagtaactaagcgtcttgaca 1135
Db 1074 ATTCCCAAAATGGAATCTGGGGAGTGTTCATCCCGAGATATTACAGCGCTTTTGACA 1133
QY 1136 gggccaacaacctcgttggcgttagctaaagaactctga 1173
Db 1134 GGGCAACAACCTCTGTGGGCTGGCCAAAGCCATCTGA 1171
RESULT 12
LOCUS E00144 1289 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding bovine calf chymosin.
ACCESSION E00144
VERSION E00144.1 GI:2168443
KEYWORDS JP 1984021392-A/1.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1289)
AUTHORS Chiyarusu,E.B.
TITLE CATTLE KIMOCIN
JOURNAL Patent: JP 1984021392-A 1 03-FEB-1984;
GENEX CORP
COMMENT OS bovine calf
PN JP 1984021392-A/1
PD 03-FEB-1984
PF 30-JUN-1983 JP 1983119481
PR 01-JUL-1982 US 82 394433. 13-APR-1983 US 83 484539 PI
CHYARUSU EI BASURETSUTO
PC C12N15/00, C07H21/04, C12N1/20, C12P19/34, C12P21/02, (C12N15/00,
C12R1:19);

Db 532 CAGAGTACTCGATACCCGTTTGGACAAACATGATGACAGGACCTGGTGGCCCAAGACC 651
Oy 656 tgtctcggtttacatgagcaagaaatggccagagagatgctcaagcttgagatattg 715
Db 652 TGTTCCTCGTTTACATGACAGAAATGGCCAGAGAACCTGCTCAAGCGGGGGCCATGC 711
Oy 716 atccatccctactacacagagatctctcactggtggtccagctcgtgagagactgagc 775
Db 712 ACCCGTCTACTACACAGAGGTCCTTCATGTTGGTCCCGTGACAGTGCAGAGTACTGGC 771
Oy 776 aattcaactgtgagacagtgtcaccatcaagcgtgtgtgtgtgtgtgtgtgtgtgtgtc 835
Db 772 AGTTCACTGTGACAGTGTACACATCAGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTG 831
Oy 836 aagctatcttgatagccggttcagctcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtc 895
Db 832 AGGCCATCTGTGACAGGAGGACCTCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 891
Oy 896 ttcagcaagctatgtgagcaacacagaaacagtaagtgatgtgtgtgtgtgtgtgtgtgtgtc 955
Db 892 TCCACAGAGCCATTTGAGACACAGAACAGTACGTTGAGTTGACATCGACTGCGACAC 951
Oy 956 accctagctacatgctcctacagttgtctttagatcaacgagcaagatgtaccactgaccc 1015
Db 952 ACTGAGCTACATGCCCCCTG 1011
Oy 1016 cctcgagctatagccagcagagatcaaggttctgacacagtgatctcaagatgtgtgtgtgtc 1075
Db 1012 CCTCGGCTATACACAGGACAGGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1071
Oy 1076 attcccaaatgtgacttctgtgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtc 1135
Db 1072 ATTCCCAAAATGTGATCTCTGGGGGATGTTTATCTCGAGAGATTTACACGCTTTTGACA 1131
Oy 1136 gggccaacaacctggtc 1173
Db 1132 GGGCCAAACACCTCGTGGGGCTG 1169

RESULT 15
LOCUS E00295 1460 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding pre-prorenin A.
ACCESSION E00295
VERSION E00295.1 GI:2168583
KEYWORDS JP 1985058077-A/3.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 1460)
Deblutsudo,B., Konarudo,U.D., Jierarudo,R.F., Arison,T.R.,
Rodaato,J.N., Jien,I.M., Donarudo,T.M. and Kurisutofuaa,G.G.
USE OF GAL YEAST PROMOTOR
Patent: JP 1985058077-A 3 04-APR-1985;
KORABORATEIBU RES INC
COMMENT
TITLE JOURNAL
AUTHORS
REFERENCE
OS Bovine
PN JP 1985058077-A/3
PD 04-APR-1985
PF 28-FEB-1984 JP 1984035472
PR 28-FEB-1983 US 83 470911
PI DEBITUSUDO BOTSUTOSUREIN, ROMARUDO UEIN DEIBISU, PI
JIERARUDO RARUFU FUINKU,
PI ARISON TAUNTUN RIGUBITI, ROBARATO JIENTORIT NOURUTON, JIEN I
MOO, PI DONARUDO TEIRAA MOA, KURISUTOFUA GOTSUDOFURIT GOFU PC
C12N15/00,C07H21/04,C07K13/00,C12N1/16,C12P21/02,C12N1/16, PC
C12R1.865),
PC (C12P21/02,C12R1.865);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;

CC *source: tissue type=Stomach mucosa;
CC *source: clone=293-207 & 293-118/37;
FH Key Location/Qualifiers
FH CDS 205..1350
FT /product='pre-prorenin A'
FT mat.peptide 205..1347
FT /product='pre-prorenin A'
FT mat.peptide 253..1347
FT /product='Prorenin'.
FEATURES
source Location/Qualifiers
1..1460
/organism='Bos taurus'
/db_xref='taxon:9913'
BASE COUNT 327 a 437 c 400 g 296 t
ORIGIN
Query Match 76.1%; Score 892.4; DB 6; Length 1460;
Best Local Similarity 87.4%; Pred. No.1.4e-260;
Matches 977; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
Oy 56 tctgtgtgttactacagctgtc 115
Db 233 TCTTCGCTCTTCCACAGGCGCTGAGATCACAGATCCCTGTGTACAAAGCAAGTCTC 292
Oy 116 tccgtgaagcgtgtgaagaaatgtgactctgaagaactcttctgtgaagaacaagatgtc 175
Db 293 TGAGGAAGCGGTGAAGGAGCATGGGCTTCTGAGAGACTTCTGACAGAAACAGCAATATG 352
Oy 176 gcatcaagcagaaagtaactccggtcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtc 235
Db 353 GCATCAGCAGCAAGTACTCTCGGCTTGCGGACAGTGGCAGCCCTGCTGACCAACTTACC 412
Oy 236 ttgatagtaataactttgtggaagatcttaccctgtgaaccccgctcaagaagtctacagctc 295
Db 413 TCGATAGTACATCTTGGAGAGATCTACCTCGGAGACCCCGCCAGGAGTTACCGTGC 472
Oy 296 tctttgatactgtgttcttctgtc 355
Db 473 TGTGTGACACTGGCTCTCTGTGACTTCTGGGTCCCTGTATCTGCAAGACCAATGTCT 532
Oy 356 gaaagacacccaagaatctgcatccgagaagctgtccacttccagaaacttaggcaaac 415
Db 533 GCAAAACCCACAGCGCTTCGACCCGAGAAATCGTCACTTCCAGAACTGTGGCAAC 592
Oy 416 cctgtctataactacactaggttagcatgtcaagaaatcttagctatgtataccgtca 475
Db 593 CCTGTCTATTCACATACGGGACAGGAGCATGCTGGGCTATGTATACGCTCA 652
Oy 476 cgtgtctcaactgtgtgacatcaacagaaagttagacttagcaccacaagaacaggtg 535
Db 653 CTGTCTCAAACTTGTGACATCGACAGACAGTGGGCTGTGACACCCAGCGCGGGG 712
Oy 536 atgtctcaactatcagaatctgcatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtc 595
Db 713 AGCTCTTACCTATTCGCAATTCGAGGGGATCTCTGGGATGTGCTTACCTCTGCTGCT 772
Oy 596 cagagtaactgataacctgtc 655
Db 773 CAGAGTACTCGATACCCGTTTGGACAAACATGATGAACAGCACTGGTGCCCAAGACC 832
Oy 656 tgttctcggtttacatgagcaagaaatggccagagatgctcaagctgtgagatattg 715
Db 833 TGTTCCTCGTTTACATGACAGAAATGGCCAGAGAACCTGCTCAAGCGGGGGCCATGC 892
Oy 716 atccatccctactacacagagatctctcactggttccagctcgtgtgtgtgtgtgtgtgtgtc 775
Db 893 ACCCGTCTACTACACAGAGGTCCTTCATGTTGGTCCCGTGACAGTGCAGAGTACTGGC 952
Oy 776 aattcaactgtgagacagtgtcaccatcaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtc 835
Db 953 AGTTCACTGTGACAGTGTACACATCAGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1012

PD 01-MAR-2001.
 PP
 PE 23-AUG-2000; 2000MO-CA00975.
 PF
 PR 23-AUG-1999; 9905-0378696.
 PS
 PA (SEMB-) SEMBIOSYS GENETICS INC.
 PP
 PV Van Rooijen G, Keon RG, Boothe J, Shen Y;
 PI
 DR WPI: 2001-226621/23.
 PP
 P-PSDE; AA000536.
 XX
 PT Producing chymosin in seeds of plants such as rice, flax, rape seed, by
 PT transforming plant cell with a nucleic acid encoding chymosin operably
 PT linked to transcription regulator and terminator sequences -
 PS
 PS Claim 9; Fig 1; 56pp; English.
 CC The sequence represents a DNA which encodes a bovine chymosin
 CC polypeptide. Chymosin can be produced in a plant seed through
 CC introduction of a chimeric nucleic acid molecule, comprising a nucleic
 CC acid sequence encoding a chymosin polypeptide operatively linked to
 CC transcription regulator and terminator sequences, into a plant cell. The
 CC sequences are useful for producing plant seeds, in particular seeds of
 CC soybean, rape seed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
 CC barley, oats, sorghum, *Arabidopsis thaliana*, potato, flax/linseed,
 CC safflower, oil palm, groundnut, Brazil nut, coconut, castor, coriander,
 CC squash, jojoba and rice.
 XQ Sequence, 1173 BP; 299 A; 308 C; 262 G; 304 T; 0 other;
 XX

	Query Match	100.0%;	Score 1173;	DB 22;	Length 1173;	
	Best Local Similarity	100.0%;	Pred. No. 0;			
	Matches 1173; Conservative	0;	Mismatches	0;	Indels	0;
OY	1 atgaacttccttaagatccttccctttacgtcttcacgcttccttggttcggcaaatctcggt	60				
Dd	1 atgaacttccttaagatccttccctttcacgcttccttggttcggcaaatctcggt	60				
OY	61 gctgttactcaogctgtgcagatacccgcattcctctctacaagaagtatctccgt	120				
Dd	61 gctgttactcaogctgtgcagatacccgcattcctctctacaagaagtatctccgt	120				
OY	121 aagcgctcgaaaggaaacatggaactcttaaagaactcttcgcaagaacaagtatgcatc	180				
Dd	121 aagcgctcgaaaggaaacatggaactcttaaagaactcttcgcaagaacaagtatgcatc	180				
OY	181 agcagcaagtaactccggttcgtagaagtctgaacgfcgccaattaccaactacttgat	240				
Dd	181 agcagcaagtaactccggttcgtagaagtctgaacgfcgccaattaccaactacttgat	240				
OY	241 agtcaatactttgggaagaatctaccctcggaaccccgcttcaagaagttaacggttctctt	300				
Dd	241 agtcaatactttgggaagaatctaccctcggaaccccgcttcaagaagttaacggttctctt	300				
OY	301 gatactggttcccttgcgaactcttggttccctctatctatctgaagaacgatgcctgcaag	360				
Dd	301 gatactggttcccttgcgaactcttggttccctctatctatctgaagaacgatgcctgcaag	360				
OY	361 aaccoccaaagaattgcataccgagaagaagtcgctccaccttcogaacttggcaaacccctg	420				
Dd	361 aaccoccaaagaattgcataccgagaagaagtcgctccaccttcogaacttggcaaacccctg	420				
OY	421 tctaatacctacgtagacagtagcaatgaagaagaaacttaaggatgataccgtaactg	480				
Dd	421 tctaatacctacgtagacagtagcaatgaagaagaaacttaaggatgataccgtaactg	480				
OY	481 tcacaacattgtggacattccaacagacagtaggaacttgaacccaagaacagtgatc	540				
Dd	481 tcacaacattgtggacattccaacagacagtaggaacttgaacccaagaacagtgatc	540				

Oy	541	tttcctcatgcagaagattcgcgtatgccttcgtgatgtgcgtaccattcgtctgcgtcaag	600
Db	541	ttccacctatgcagaatttcgcgtgcaccccttggtatgtgataaccacgcgtctgcgtcaag	600
Oy	601	tactcgtataacctgtgttttgacaacatgatgatbaaccgacactagtagtcltaagaactgttc	660
Db	601	tattcgcgtataacctgtgttttgacaacatgatgatbaaccgacactagtagtcltaagaactgttc	660
Oy	661	tcggtttacatgatgcagaagaatggtccagagagcattgctcaagctctggagctattgatcca	720
Db	661	tcggtttacatgatgcagaagaatggtccagagagcattgctcaagctctggagctattgatcca	720
Oy	721	ttccactacaaaggaattcttctcacttggttcacgtacacgttgtagagagatcgtggaatttc	780
Db	721	ttccactacaaaggaattcttctcacttggttcacgtacacgttgtagagagatcgtggaatttc	780
Oy	781	actgtgacacagtgtgcacacatccacgcgtgtgtgttgcattggaaggtgatgtcaagct	840
Db	781	actgtgacacagtgtgcacacatccacgcgtgtgtgttgcattggaaggtgatgtcaagct	840
Oy	841	attcttggataacggttacgltccaaagtgtgtgcgacctagacgcgacattctcaacatcaag	900
Db	841	attcttggataacggttacgltccaaagtgtgtgcgacctagacgcgacattctcaacatcaag	900
Oy	901	caaacctattggagccacacagaacacgcgtacggtgagtttggatatgattgcgacaactt	960
Db	901	caaacctattggagccacacagaacacgcgtacggtgagtttggatatgattgcgacaactt	960
Oy	961	agctacatgccttaacagttctgtcttcttgagatccaacgacgaagatgtaccactatgaccccttc	1020
Db	961	agctacatgccttaacagttctgtcttcttgagatccaacgacgaagatgtaccactatgaccccttc	1020
Oy	1021	gctctataccagccagatccaaggttcttcgacacagtgtatccaaggttgagaacacattcc	1080
Db	1021	gctctataccagccagatccaaggttcttcgacacagtgtatccaaggttgagaacacattcc	1080
Oy	1081	cagaaatggatcttggagaagtgtgttcattctcgtgagttacttaacgcgtcttlttgaacgggcc	1140
Db	1081	cagaaatggatcttggagaagtgtgttcattctcgtgagttacttaacgcgtcttlttgaacgggcc	1140
Oy	1141	aacaacctcgtttggcgtactgaataagcaattctga	1173
Db	1141	aacaacctcgtttggcgtactgaataagcaattctga	1173
RESULT 2			
ID	AAS00570 standard; DNA: 3957 BP.		
XX	AAS00570;		
XX	14-MAY-2001 (first entry)		
DE	Bovine phaseolin promoter pre-pro-chymosin-phaseolin terminator.		
XX	Chymosin: transcription regulator; terminator sequence: soybean; corn;		
KW	pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;		
KW	barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;		
KW	safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;		
KW	squash; jojoba; ds; phaseolin; promoter; terminator; mutant; French bean;		
XX	Chimeric - Bos sp.		
OS	Chimeric - Phaseolus vulgaris.		
XX	Key		
XX	promoter		
FT	1..1553		
FT	/*tag- a		
FT	/note= "Phaseolin promoter"		
FT	1554..2726		
FT	/*tag- b		
FT	/product= "Bovine pre-pro-chymosin"		
FT	2727..3957		
FT	terminator		

```

FT      /tag= C
FX      /note= "Phaseolin terminator"
PN      WO200114571-A1.
XX
XX
XX      01-MAR-2001.
PD
XX
XX      23-AUG-2000; 2000WO-CA00975.
PE
XX
XX      23-AUG-1999; 99US-0378696.
PR
XX
XX      (SEMB-) SEMBIOSYS GENETICS INC.
PA
XX
XX      Van Rooijen G, Keon RG, Boothe J, Shen Y;
PI
XX      WPI; 2001-226621/23.
DR      P-PSDB; AAU00536.
XX
XX      Producing chymosin in seeds of plants such as rice, flax, rape seed, by
PT      transforming plant cell with a nucleic acid encoding chymosin operably
PT      linked to transcription regulator and terminator sequences -
XX
XX      Example 1; Fig 2; 56pp; English.
PS

```

CC The sequence represents a chimeric polynucleotide comprising a
CC pre-pro-chymosin, a phaseolin promoter and a phaseolin terminator.
CC Chymosin can be produced in a plant seed through introduction of a
CC chimeric nucleic acid molecule, comprising a nucleic acid sequence
CC encoding a chymosin polypeptide operatively linked to transcription
CC regulator and terminator sequences, into a plant cell. The sequences are
CC useful for producing plant seeds, in particular seeds of soybean, rape
CC seed, sunflower, cotton, corn, tobacco, alfalfa, barley, oats,
CC sorghum, *Arbidiopsis thaliana*, potato, flax/linseed, safflower, oil palm,
CC groundnut, Brazil nut, coconut, castor, coriander, squash, jojoba and
CC rice.
XX
XX Sequence 3957 BP; 1263 A; 790 C; 609 G; 1295 T; 0 other:
SQ

Query Match	100.0%;	Score 1173;	DB 22;	Length 3957;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1173; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

OY	1	atgaactctcttaagttctcttcctcttctaagctcttccttctgttctggfcaactcgtt	60
Db	1554	atgaactctcttaagttctcttcctcttctaagctcttccttctgttctggfcaactcgtt	1613
OY	61	gctttactacggtgtgtgagatcacccgattccctcttaaaagftraagctcttcgt	120
Db	1614	gctttactacggtgtgtgagatcacccgattccctcttaaaagftraagctcttcgt	1673
OY	121	aagcgctggaagagacatctgactctagaagactcttgcagaacaacagatgacac	180
Db	1674	aagcgctggaagagacatctgactctagaagactcttgcagaacaacagatgacac	1733
OY	181	agcagcaagtaactccggtcttcggttgaagtgtcagtcggtgcacttaccactacttgat	240
Db	1734	agcagcaagtaactccggtcttcggttgaagtgtcagtcggtgcacttaccactacttgat	1793
OY	241	agtaataactttggtgaagatctacctcccgaaaccccgctcaagatcaccgttctctt	300
Db	1794	agtaataactttggtgaagatctacctcccgaaaccccgctcaagatcaccgttctctt	1853
OY	301	gatactgttccctcttgactcttggttccctctatctatctgaagaagaaatgctgcgaag	360
Db	1854	gatactgttccctcttgactcttggttccctctatctatctgaagaagaaatgctgcgaag	1913
OY	361	aaccaccaaaagattcgatccgagaagaatgctccaccttcacagaacttgaagcaaaccttg	420
Db	1914	aaccaccaaaagattcgatccgagaagaatgctccaccttcacagaacttgaagcaaaccttg	1973
OY	421	tctatacactacggttaagaagtaagatcgaagaagatcttgaatgatataccgtcacctc	480

Dd	1974	tcctacaciacggtlacaggtacgtacgtcaaggaatcttaagctatgataccgtcactgtc	2033
Qy	481	tcacaactgtgagcatctcaacagacagactagagcttatagaccacaagaacaggtatctc	540
Dd	2034	tcacaactgtgagcatctcaacagacagactagagcttatagaccacaagaacagaggtatctc	2093
Qy	541	ttcacctatgcagaatctcgaatggcatcctctgttatgtgacataccatacgtctcggtcagag	600
Dd	2094	ttcacctatgcagaatctcgaatggcatcctctgttatgtgacataccatacgtctcggtcagag	2153
Qy	601	taccgatacctgtgtttgtgacaacatgatagaaacgcacactagctagctccaagactgttc	660
Dd	2154	taccgatacctgtgtttgtgacaacatgatagaaacgcacactagctagctccaagactgttc	2213
Qy	661	tcgatttatactgacaaggaatgtgcacagggagaaatgactcaagctctgtgagcatgtatcca	720
Dd	2214	tcgatttatactgacaaggaatgtgcacagggagaaatgactcaagctctgtgagcatgtatcca	2273
Qy	721	tcctactataccaagatctctctcacgtggttccaaagtcactgtgcaagtaactgagcaattc	780
Dd	2274	tcctactataccaagatctctctcacgtggttccaaagtcactgtgcaagtaactgagcaattc	2333
Qy	781	actgtgacaagtgctcaaccatacagcggctgtgtgtttgtgcactgtggaaggtggatgtccaagct	840
Dd	2334	actgtgtgacaagtgctcaaccatacagcggctgtgtgtttgtgcactgtggaaggtggatgtccaagct	2393
Qy	841	atctctgataccggtatgctcaacagctcgggtccggaacctatagaagagatcttcaaatctag	900
Dd	2394	atctctgataccggtatgctcaacagctcgggtccggaacctatagaagagatcttcaaatctag	2453
Qy	901	caagactatgtgagaccacaacagaacccagtaacggttgagcttgacataagatctgacaaacctt	960
Dd	2454	caagactatgtgagaccacaacagaacccagtaacggttgagcttgacataagatctgacaaacctt	2513
Qy	961	agctacaatgacctacagattgtctcttgatatacaagcgcaagaatgtataccacvtacccctctc	1020
Dd	2514	agctacaatgacctacagattgtctcttgatatacaagcgcaagaatgtataccacvtacccctctc	2573
Qy	1021	ggctatacccgccaagatcaagggctcttcacacatgtgatacttccaagtgagaaacatcttc	1080
Dd	2574	ggctatacccgccaagatcaagggctcttcacacatgtgatacttccaagtgagaaacatcttc	2633
Qy	1081	cagaatatyatactctgagagagatgtgtctcaatctgttgatactatacagcgtcttggacaagggcc	1140
Dd	2634	cagaatatyatactctgagagagatgtgtctcaatctgttgatactatacagcgtcttggacaagggcc	2693
Qy	1141	aacaaactcgttggctagctaaagcaatctga 1173	
Dd	2694	aacaaactcgttggctagctaaagcaatctga 2726	

RESULT	3	
AA91157		
ID	AA91157	standard; DNA; 1291 BP.
XX		
AC	AA91157;	
XX		
DT	07-JUN-1990	(first entry)
XX		
DE	Cloned sequence of (pro)chymosin.	
XX		
KW	Lactic acid bacteria; cheese; Streptococcus cremoris SK112; chymosin;	
KW	protease; PSK112; ss.	
XX		
FH	Key	Location/Qualifiers
FT	precursor_RNA	29..1123
FT		/*tag= a
FT		/product=; prochymosin
XX		
PN	NE8701378-A.	
XX		
PD	02-JAN-1989.	
XX		

PF 12-JUN-1987; 87NL-0001378.
XX
XX 12-JUN-1987; 87NL-0001378.
XX
XX (NEZU-) NEDERL. INS ZUIVELON.
XX
XX Simons AFM, De Vos WM;
PI
XX
XX WPI: 1989-030097/04.
DR P-PSDB; AAP94144.
PT DNA fragment having region specific for lactic acid bacteria -
PT is contained in plasmid in microorganism used in prodn. of
PT protein and food prodn. eg cheese.
XX
XX
PS Disclosure; fig 2; 43pp; Dutch.
XX
XX The DNA encoding prochymosin can be cloned into a plasmid (esp. from
CC S. cremoris SK112) and used to produce large amts of the protein by
CC recombinant DNA techniques. This could overcome the shortage of
CC prochymosin due to shortage of calf stomachs and increasing cheese
CC prodn. Prochymosin is also used in prodn. of yoghurt, butter and
CC buttermilk. See also AAN91158-N91160.
XX
XX
SQ Sequence 1291 BP; 323 A; 385 C; 326 G; 257 T; 0 other;

Query Match 76.6%; Score 898.8; DB 10; Length 1291;
Best Local Similarity 87.7%; Pred. No. 2; 3e-273;
Matches 981; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 56 tctgtgctgtacacagcgtcgtgagatcacccgcattccctctctacaaagtaagctc 115
DB 9 tcttcgctctccacagcggtgagatcacccgcattccctctctacaaagtaagctc 68
QY 116 tccgtaagcgctgaaagaaatgactcttagaagctctctgcaaaacaagatc 175
DB 69 tgaagaagcgctgaaagaaatgactctgagagactctctgcaaaacaagatc 128
QY 176 gcatacagcaagaatctcggctcgtgtaagtgtgtagagcttccactaacactacc 235
DB 129 gcaacagcaagaatctcggctcgtggaagtgccagcgctccctgaccactacc 188
QY 236 tctgatacatatactctgggaagaatctacactcggaaccgccccaagaagttcacctc 295
DB 189 tggtagtagtaactcttgggaagaatctacactcggaaccgccccaagaagttcacctc 248
QY 296 tcttgtaactgtctcctcgtactctgtggtccctctatctactgcaagaagatgct 355
DB 249 tgttgtaactgtctcctcgtactctgtggtccctctatctactgcaagaagatgct 308
QY 356 gcaagaacacccaagaatctgatacgcgaagaagtgctacacttccagaactaggcaaac 415
DB 309 gcaaaacacccaagcttcgacccggaagaagtcgctccactccacactccagactggcaagc 368
QY 416 cctgtctatacacaatgacagtgacagtgacatgcaagaatcttagagctatgatacgtca 475
DB 369 cccgtctatccactaagcagacgacgacatgcaagcgcctctgcatgacacgtca 428
QY 476 ctgtctccaacatgtggaacatccaacagatagaaactagaccaccaagaccagt 535
DB 429 ctgtctccaacatgtggaacatccaacagatagaaactagaccaccaagaccagt 488
QY 536 atgtctccaactagcagaatgcatgcatcctgtgtatgcatcaccactgctgct 595
DB 489 acgtctccaactagcagaatgcatcgaacggtcctgggaatgcttaccaccccgctgct 548
QY 596 cagaagctacgatacgtgtgttgaacaacatgatgaacccgacactagtaagtaagact 655
DB 549 cagaagctacgatacgtgtgttgaacaacatgatgaacccgacactggtgccaagacc 608
QY 656 tgtctcgtgttatacgtgacaagaatgcccagagaagcatgtccaagcttgagactatg 715

DB 609 tgtctcgtgttatacgtgacaagaatgcccagagaagacatgctcaccgtggtggccatcg 668
QY 716 atccatcctactacaagagatctcttcactgggttccagactgctgcaagtaactgac 775
DB 669 acccgtctactacaagagatctcttcactgggttccagactgctgcaagtaactgac 728
QY 776 aattcactgtgacagtgatccatccatcagcgtgtgtgtgttctgatactgaagtgatg 835
DB 729 agttcactgtgacagtgatccatccatcagcgtgtgtgtgtgtgtgtgagcgtgctgc 788
QY 836 aagctatcttgataccggtgacgttccaaagtgtgtgacttagcagagatctcaca 895
DB 789 aggcataccttgataccggtgacgttccaaagtgtgtgtgtgtgtgtgtgtgtgtgtgt 848
QY 896 ttcaagcagctattggagccacacagaccagtaacgtgtgattgataattgacgaca 955
DB 849 tccagcagcgcatgtgagccacacagaccagtaacgtgtgattgataattgacgaca 908
QY 956 acctgaagctatgctcctacagttgtctttagatcaacggcgaagatgaccactgaccc 1015
DB 909 acctgaagctatgctcctacagttgtctttagatcaacggcgaagatgaccactgaccc 968
QY 1016 cctccgctataccagccaggaatcaaggtctcgcacacagtgatctcgaagtgagaacc 1075
DB 969 cctccgctataccagccaggaatcaaggtctcgcacacagtgatctcgaagtgagaatc 1028
QY 1076 attccagaatgatacttggagatgtgtcattcgtgtgagtaacagcgtcttgaca 1135
DB 1029 attccagaatgatacttggagatgtgtcattcgtgtgagtaacagcgtcttgaca 1088
QY 1136 gggccaacaactcgtgtggtgtagtaagaacatcga 1173
DB 1089 gggccaacaactcgtgtggtgtagtaagaacatcga 1126

RESULT 4
AAN40295
ID AAN40295 standard; mRNA; 1175 BP.
XX
XX AAN40295;
AC
XX
XX 04-FEB-1992 (first entry)
DE
XX
XX Sequence encoding a polypeptide displaying milk clotting activity.
KW Cheese-making; recombinant protein; rennet substitute; milk clot; ss.
XX
OS Bos taurus.
XX
XX
FH Key
FH sig_peptide
FT /*tag- a
FT mat_peptide 70..1166
FT /*tag- b
XX
XX
XX EPI23928-A.
XX
XX
XX 07-NOV-1984.
PD
XX
XX
XX 30-MAR-1984; 84EP-0103551.
PF
XX
XX
XX 31-MAR-1983; 83US-0480860.
PR
XX
XX (CODO-) CODON GENETIC ENG.
PA
XX
XX
XX Cashion LM, McCaman MT, Rice CW, Sias SR;
XX
XX
XX WPI: 1984-277277/45.
DR P-PSDB; AAP40559.
XX
XX
XX Recombinant DNA coding for milk clotting polypeptide - which is
PT expressed in transformed bacteria
XX

PS Claim 6; Fig 2; 39pp; English.

CC Also claimed is E.coli JM83/PLC7 (ATCC 39325) which is transformed
CC with PLC7 contg. the prorenin derived sequence fused in phase with
CC B-galactosidase. The PLC7 prorenin expression plasmid includes
CC Sequences which code for both the pseudorenin and mature rennin
CC cleavage sites between AAs 28-29 and AAs 42-43, respectively.
XX

Sequence 1175 BP; 263 A; 355 C; 319 G; 238 U; 0 other;

Query Match 76.5%; Score 897.2; DB 5; Length 1175;
Best Local Similarity 68.2%; Pred. No. 7e-273;
Matches 763; Conservative 217; Mismatches 138; Indels 0; Gaps 0;

OY 56 tcgttcgcttaccacgctctgagatccacccgctcctcctacagaagtgaatctc 115
DB 49 uuuuugcucucucccaaggcgccgagaucaucccaagucuuacaagaagcaguc 108
OY 116 tccglaaggcgctgaaggacatgtgactctagaagactcttgtagaacaacagatg 175
DB 109 ugaggaaagcgcuagaagcagcuuucugagacucucugagaagaacagcaguanug 168
OY 176 gcatcagcagaagtaactcgcgcttggtgaagtgtcagcgttcgacttaccactacc 235
DB 169 gcaucagcagaagacucgcgcuucgggagugugccagcgugcccucaacuaacc 228
OY 236 ttgtagtcaatcactcttggaagatctaccctcggaacccgctcacaagtccagctc 295
DB 229 uggaagaagcagacucuuugagagaucacucggagaccgcccagagagauccagc 288
OY 296 tcttgatcactggtcctcctcactctctggtccctcctatcactcagaagcagatgct 355
DB 289 uguuugacacugcucucucugacucucuguaaccucuaucucgcaagcagaugccu 348
OY 356 gcaagaacaccacaagatcgatccgagaagaatcgctccaccttcagaactaggcaaac 415
DB 349 gcaaaaaaccacagcgcuucgacccgagaaagucgucacacucacagaccugcgaaagc 408
OY 416 ccttgctatcactcagatcagatgacatgacagaagaaatcttgatgatactacgta 475
DB 409 cccugucuaucacacagcgagacagcaugcagaggaagcauccuacugacacccguc 468
OY 476 ctgtctccaaactgtgtgacatctcaacagacagtaagagacttagccccaagaacagtg 535
DB 469 cugucuccaacaauugugacacacagagacaguanagccuagaccacagagaccgug 528
OY 536 atgtcctcactatcgaaatcgatgcacccctgtgtatgcataccacatcgctcgct 595
DB 529 agucucuaucacuaugccgaauucgagcgaucucugggagugccuacccucugcuc 588
OY 596 cagagttactcgatcactcgtgtttgacaacatgataagaccgactatgactcaagact 655
DB 589 cagaguaucagaccccgguuuuugacaacauagaacagagcugugugcccaagacc 648
OY 656 tgttctcgcttaccatcagcaggaatggtccagagagacatcctcagcttgagactatg 715
DB 649 uguuucuguuuacacugacaggaugccagggagagcagucacgucuggggagccacug 708
OY 716 atccatcactacacagagatccttcttcctcgtgtcctcagctactgtgacagatctgc 775
DB 709 acccgucuaacacacagcagucucugcugugcccgagacagugacagcagucagucgc 768
OY 776 aattcactgctgacagtgctacacacagcgtgtgtgtgtgacgtgaagtgtgacttc 835
DB 769 aguuacacugugagacagucacacacagcggugugugugccugugagugugcuguc 828
OY 836 aagctatcttgataccgctacgctcgaagctggtcgcagctagcagcagacattccaca 895
DB 829 aggcacacucugagacagcgacacucacagcugugcgccacagcagcagacacacaca 888
OY 896 ttccagcaagctattggagcacaacagaaaccagtaactgtgatttgacataatgtcgaca 955

DB 889 uccagcagcgcauugggagccacacagaaaccaguaacgauugauuugacucagucgaca 948
OY 956 accctagctacatcagctcagttgtcttgagatcaaggcgaagatgtaccacagacc 1015
DB 949 accugagcuaacugccacuguguuuugauaauagaagcaaaauagaaccacagacc 1008
OY 1016 ccttcgctatccagcaggaagtggttcgtgacagtgatctcagagtgagaaacc 1075
DB 1009 ccuucgcuuaacacagcaagcagcgcuucuguaaccagugcucacagagugaaauac 1068
OY 1076 attccagaatgagatcttgaggagatgtgttcattcgtgtgactacaagctcttgaca 1135
DB 1069 auuccaagaauagauccugggagugauuuaucaggagagauuaucagcgucuuuagaca 1128
OY 1136 gggccaacacctgttggtgactgaagaacatctga 1173
DB 1129 gggccaacaacccuggggagcgccaaagcaacucaga 1166

RESULT 5

AAT03006
ID AAT03006 standard; DNA; 1240 BP.

AAT03006;

13-JUN-1996 (first entry)

Chymosin open reading frame.

al-3; albino mutant; light-regulated; Neurospora; bread mould;
heterologous gene; expression; control; chymosin; ss.

Mammalian sp.

W09530739-Al.

16-NOV-1995.

09-MAY-1995; 95WO-US05716.

10-MAY-1994; 94US-0240372.

(UYHA-) UNIV HAWAII.

Kato EK, Stuart WD;

WPI; 1995-404108/51.

Nucleic acid for expression of heterologous protein - contains
albino promoter for light induced expression in filamentous fungi

Example 2; Fig 7; 26pp; English.

The DNA is that of a mammalian gene (open reading frame) encoding
chymosin. The gene was placed in operable linkage with the al-3
promoter (see AAT03005). The al-3 gene controls the production of
geranyl geranyl pyrophosphatase (GGPP) synthetase. GGPP is a precursor
for carotenoids and xanthophylls. It has been shown that exposure to
light increases the transcription level of GGPP synthetase 15-45 fold.
This activates a number of genes in the common bread mould, Neurospora.
This can be used to regulate the expression of genes encoding
heterologous proteins, e.g. chymosin, in recombinant production systems.
Use of a light-regulated promoter is a simple and effective way to
control expression and allows timing to be adapted to the physiological
status of the host.

Sequence 1240 BP; 274 A; 374 C; 339 G; 253 T; 0 other;

Query Match 76.5%; Score 897.2; DB 16; Length 1240;
Best Local Similarity 87.7%; Pred. No. 7.2e-273;
Matches 980; Conservative 0; Mismatches 138; Indels 0; Gaps 0;


```

Db 313 gcaaaaaccaccgcttcgaccggaagaatgcgtccacctccagaacctgggcaagc 372
OY 416 cctgtctatactacagtagcagtagatgcaaggatcttagcatatgataccgtca 475
   || ||||| ||||| ||||| ||||| || ||||| ||||| |||||
Db 373 cctgtctataccctacagcagcagcagcagcagcagcagcagcagcagcagcagc 432
OY 476 ctgtctccaacattgttgacaattcaacagacagtaggaacttagcaccgaagaacagtg 535
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 433 ctgtctccaacattgttgacaattcaacagacagtaggaacttagcaccgaagaacagtg 492
OY 536 atgtctccaacattgttgacaattcaacagacagtaggaacttagcaccgaagaacagtg 595
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 493 acgtctccaacattgttgacaattcaacagacagtaggaacttagcaccgaagaacagtg 552
OY 596 cagatctacgataccgtgttctacacattatgaaccgacacccctgatacgaagact 655
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 553 cagatctacgataccgtgttctacacattatgaaccgacacccctgatacgaagact 612
OY 656 tgtctcgtttacatagagagaatgccaagagagacatgctcaagcttgagagctattg 715
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 613 tgtctcgtttacatagagagaatgccaagagagacatgctcaagcttgagagctattg 672
OY 716 atccatctactataccagcagatctcttctactggtgtccagatcgtgacagagactggc 775
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 673 accgctctactataccagcagcagcagcagcagcagcagcagcagcagcagcagc 732
OY 776 aattcactgtgagacagctcaccatcagcagctgtgtgtgtgtgtgtgtgtgtgtgtgt 835
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 733 agttcactgtgagacagctcaccatcagcagcagcagcagcagcagcagcagcagcagc 792
OY 836 aagctactcttgagataccggtatcctcaagctgtgtcagcagctagcagcagcagcagc 895
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 793 agcgcatacttgagacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 852
OY 896 ttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 955
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 853 tccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 912
OY 956 accttaagctacatagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1015
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 913 acctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 972
OY 1016 cctcgcctataccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1075
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 973 cctcgcctataccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1032
OY 1076 attccagaagaatgagctctgagagagatgtgtcattcgtgagatcactacagcagcagc 1135
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1033 attccagaagaatgagctctgagagagatgtgtcattcgtgagatcactacagcagcagc 1092
OY 1136 ggcccaacaacctcgttgagcagcagcagcagcagcagcagcagcagcagcagcagc 1173
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1093 ggcccaacaacctcgttgagcagcagcagcagcagcagcagcagcagcagcagcagc 1130

```

```

RESULT 8
AAN20043
ID AAN20043 standard; DNA; 1460 BP.
XX
AC AAN20043;
XX
XX 16-DEC-1992 (first entry)
XX
XX DE Pre-prorennin-A gene DNA sequence.
XX
XX KW Pre-pro-rennin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;
XX KM protease; milk-clotting enzyme; ss.
XX
XX Bos taurus.
XX
XX Key Location/Qualifiers
XX CDS 205..1350
FT /*tag= a

```

```

FT FT /label=
FT PN /note= "pre-prorennin-A gene"
XX GB2091271-A.
XX
XX 28-JUL-1982.
PD
XX 15-JAN-1982; 82GB-0001120.
PF
XX 01-DEC-1981; 81US-0325481.
PR 16-JAN-1981; 81US-0225717.
XX
XX (COLB ) COLLABORATIVE RES INC.
XX
XX Alford BL, Mao J, Moir DT;
XX
XX WPI: 1982-62028E/30 (62028E).
DR P-PSDB; AAP20038.
DR
XX
XX Transformed cells producing rennin and its precursors - contg.
PT appropriate recombinant DNA material
XX
XX PS Disclosure; Table 1; 39pp; English.
XX
XX Bases 1-204 and 1351 to 1460 are attached to the pre-prorennin but
CC can be removed and are not essential to use of the gene in
CC expression. The gene may be ligated into plasmid pCDE21 and
CC expressed in E. coli. The resulting expressed enzyme is a well
CC known milk-clotting enzyme used in cheese-making.
XX
XX SO Sequence 1460 BP; 329 A; 440 C; 397 G; 294 T; 0 other;

```

```

Query Match 76.4%; Score 895.6; DB 3; Length 1460;
Best Local Similarity 87.6%; Pred. No. 2.5e-272;
Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
OY 56 tctgtgtgttactcagcgtctgagatcaccgcagatccctctacaagaagtagtctc 115
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 233 tcttgccctctcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 292
OY 116 tccgtaagcgcgtgaagaacatgactcttagagagactcttgtagaagaacagatga 175
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 293 tgaagaagcgcgtgaagaacatgactcttagagagactcttgtagaagaacagatga 352
OY 176 gcatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 235
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 353 gcatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 412
OY 236 ttgatagtcaactcttggaagatctactccggaaccgcgcctcaagaagttcacgcgtc 295
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 413 tgaatagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 472
OY 296 tcttgatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 355
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 473 tgttgacacatggtccctcctgactctcgtgttaacctctatctactcagcagcagcagcagc 532
OY 356 gcaagaacacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 415
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 533 gcaaaacacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 592
OY 416 cctgtctatacactacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 475
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 593 cctgtctatacactacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 652
OY 476 ctgtctcaaaattgtgacattcaacagacagcagcagcagcagcagcagcagcagcagcagc 535
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 653 ctgtctcaaaattgtgacattcaacagacagcagcagcagcagcagcagcagcagcagcagc 712
OY 536 atgtctcaactatcagaattcgaatgcatcctgtgtatgagcagcagcagcagcagcagcagc 595
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 713 acgtctcaactatcgaattcgcagcagcagcagcagcagcagcagcagcagcagcagcagc 772

```

```

QY 596 cagagctactcgataccgtgtgtttgacaacatgataaccgacactagctcaagact 655
Db 773 cagagctactcgataccgtgtgtttgacaacatgataaccgacactagctgtgcccagacc 832
QY 656 tgttctgtgttacttgacaggaatgagccagagagacatgctcagctgtgagactatcg 715
Db 833 tgttctgtgttacttgacaggaatgagccagagagacatgctcagctgtgagactatcg 892
QY 716 atccatctactacacagatctcttcctggtgttcacgctcctgttgacagcgttctgac 775
Db 893 accgctctactacacagatctcttcctggtgttcacgctcctgttgacagcgttctgac 952
QY 776 aattcactgtgacagctgttcacacatcagcgtgtgtgtgttcagctgaaagtgtgac 835
Db 953 agttcactgtgacagctgttcacacatcagcgtgtgtgtgttcagctgaaagtgtgac 1012
QY 836 aagctactcttgatacgtgtacgttcacagctgtgttcagacatgacagcagatctcaca 895
Db 1013 aggcacatctgtgacacagcgttcacacagctgtgttcagacagcagcagatctcaca 1072
QY 896 ttccagcaagctatgtgagccacacagaaacagtaagcgtgtgagtttaccatagatttcgaca 955
Db 1073 tccagcaagcctatgtgagccacacagaaacagtaagcgtgtgagtttaccatagatttcgaca 1132
QY 956 acctagctactacgtcctacagctgtgtcttggatcacaagcgaagatgtaaccacagacc 1015
Db 1133 acctgagctactacgtcctacagctgtgtcttggatcacaagcgaagatgtaaccacagacc 1192
QY 1016 cctccgctatatacagcagcagatcaagggtctctgcacacagtgatcagagtgagaacc 1075
Db 1193 cctccgctatatacagcagcagatcaagggtctctgcacacagtgatcagagtgagaacc 1252
QY 1076 attcccaagaatgtgattctggagatgtgttcacgttcgtgagagctacagagcgtcttcgaca 1135
Db 1253 attcccaagaatgtgattctggagatgtgttcacgttcgtgagagctacagagcgtcttcgaca 1312
QY 1136 gggccaacaacactcgttggcgtgagtaagaacatctga 1173
Db 1313 gggccaacaacactcgttggcgtgagtaagaacatctga 1350

```

RESULT 9
AAZ06463 standard; DNA; 2733 BP.

```

AC AAZ06463:
XX
XX
XX 29-NOV-1999 (first entry)
XX
XX
XX 2.7 Kbp HindIII fragment of pSBSOTPMT (oleosin-chymosin fusion gene).
XX
XX oil-body; lipid body; oleosome; spherosome; separation;
XX fusion protein; heterologous polypeptide; commercial production;
XX plasmid; ds.
XX
XX Synthetic.
XX
XX
XX Key Location/Qualifiers
XX CDS 850..2729
XX FT /*tag= a
XX FT /product= "oleosin-spacer-Met-prochymosin"
XX FT /note= "CDS contains an intron"
XX FT 850..1202
XX FT /number= 1
XX FT 1203..1442
XX FT /number= 1
XX FT 1443..2733
XX FT /number= 2
XX
XX US5948682-A.
XX
XX 07-SEP-1999.
XX

```

```

PF 25-APR-1997; 97US-0846021.
XX
XX 25-APR-1997; 97US-0846021.
PR 22-FEB-1991; 91US-0659835.
PR 16-NOV-1993; 93US-0142418.
PR 30-DEC-1994; 94US-0366783.
XX
XX (SEMB-) SEMBIOSYS GENETICS INC.
XX
XX MoJoney MM:
XX
XX WPI; 1999-517960/43.
XX P-PSDB; AAV33830.
XX
XX Expression of a heterologous polypeptide on an oil body protein is
XX useful for the production of e.g. enzymes, antibodies, hormones
XX
XX Claim 15; Fig 6; 48pp; English.
XX
XX This is the nucleotide sequence of a HindIII fragment containing the
XX oleosin-spacer-Met-prochymosin sequence. This HindIII fragment was
XX joined to a nopaline synthase terminator and cloned into binary vector
XX pCGN1559. The resulting plasmid was called pSBSOTPMT and introduced
XX into A. tumefaciens. The resulting bacterial strain was used to transform
XX B. napus plants.
XX
XX The DNA which encodes a chimeric fusion protein that consists of the
XX oil-body targeting sequence, a transcription regulation sequence and the
XX DNA of the protein of interest can be used to produce antibodies,
XX glycanases, hormones, proteases, protease inhibitors, seed storage
XX proteins, thrombin inhibitors, hirudin, interleukins, chymosin, cystatin,
XX xylanase, carp growth hormone, zein or a collagenase.
XX
XX The enzyme may be cleaved from the oil body protein or used in
XX association with the oil body fraction.
XX
XX Allows production of commercially important proteins on a superior scale
XX to production by conventional systems. The expressed heterologous
XX protein can be easily separated from host cell components
XX due to the use of the oil body as a carrier protein.
XX
XX Sequence 2733 BP; 699 A; 681 C; 684 G; 669 T; 0 other:
XX

```

Query Match 76.3%; Score 894.8; DB 20; Length 2733;
Best Local Similarity 88.4%; Pred. No. 6.3e-272;
Matches 971; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```

QY 76 gctgagatccgccgactctctcttcaaaagtaagtctctcgtgaagcgtgaagaa 135
Db 1630 gctgagatccgccgactctctcttcaaaagtaagtctctcgtgaagcgtgaagaa 1689
QY 136 catgactcttctgaagactctcttgagaacacagatgagatcagcagcaagctacc 195
Db 1690 catgactcttctgaagactctcttgagaacacagatgagatcagcagcaagctacc 1749
QY 196 ggtctcgttgaagctgtagcgtgacacttaccactacactatgataagcaacttgg 255
Db 1750 ggtctcgttgaagctgtagcgtgacacttaccactacactatgataagcaacttgg 1809
QY 1750 ggtctcgttgaagctgtagcgtgacacttaccactacactatgataagcaacttgg 1809
QY 256 aagatctactcgtgaaccccgccctcgaagatcagcttctcttgaatactggtctct 315
Db 1810 aagatctactcgtgaaccccgccctcgaagatcagcttctcttgaatactggtctct 1869
QY 316 gactcttggttccctctatctactgcaagagcaatgctgtaagaacacccaagatc 375
Db 1870 gactcttggttccctctatctactgcaagagcaatgctgtaagaacacccaagatc 1929
QY 376 gatccgagaagctgtccacttccagaaacttaggcaaaccttgcataatacctagct 435
Db 1930 gatccgagaagctgtccacttccagaaacttaggcaaaccttgcataatacctagct 1989
QY 436 acaggtagcatgcaagaatcttaggtatgatacagctcactgcttccaacatgttgac 495
Db 1990 acaggtagcatgcaagaatcttaggtatgatacagctcactgcttccaacatgttgac 2049

```

OY	496	atctaacagacagtagtaggacttaagcaccaccaagaaccaagtgatgcttcttaactatgacga	555
Db	2050	atccacagacagcagtaggctctgaacaccacagagaccgcggagacgcttcttaactatgccga	2109
OY	556	ttcgatgcatcctctgtatgtagcatcccatcgcctgcgctgaagtagtactgcataccttg	615
Db	2110	ttcgacgcggagaccgcggagatgagcctccctccctcgcctccacagatgactcgatataccg	2169
OY	616	tttagacaacatgataaaccgacacactagtagtactcaagaattgttcctggtttatacagac	675
Db	2170	tttagacaacatgataaaccgacacactagtagtactcaagaattgttcctggtttatacagac	2229
OY	676	aggaaatgccaagagagacatgcttccacgcgtctgtagcgtatgtatcatcctactatacaga	735
Db	2230	aggaaatgccaagagagacatgcttccacgcgtctgtagcgtatgtatcatcgtgagacggttc	2289
OY	736	tctcttaactggtgttccatgacatctgtagcagcagtagtagcgaattcaactgtagacggttc	795
Db	2290	tccctgacatcggtgggccggtgagacgtagcagcagtagtagcgaattcaactgtagacggttc	2349
OY	796	accatacagcgt	855
Db	2350	accatacagcgt	2409
OY	856	acgttccaaagctgtgtcggacctaagcagcagcatcttcaacatccagcaagctatgttaggc	915
Db	2410	acgttccaaagctgtgtcggacctaagcagcagcatcttcaacatccagcaagctatgttaggc	2469
OY	916	acacagaacccaagtagcgt	975
Db	2470	acacagaacccaagtagcgt	2529
OY	976	gttgcctttgagatcaacacggcagaagatgtaaccacatcccccctcgcctatacagcagcag	10355
Db	2530	gttgcctttgagatcaacacggcagaagatgtaaccacatcccccctcgcctatacagcagcag	2589
OY	1036	gatacaagggtctgtcaccaagtgatctccagaagtgagaaacccattccacgaatgtgacttg	10955
Db	2590	gatacaagggtctgtcaccaagtgatctccagaagtgagaaacccattccacgaatgtgacttg	2649
OY	1096	ggagatgtgttcatctgt	11555
Db	2650	ggagatgtgttcatctgt	2709
OY	1156	ctagctaaagaacatctga 1173	
Db	2710	ctagctaaagaacatctga 2727.	
RESULT 10			
AA014051			
ID	AA014051	standard; DNA; 1210 BP.	
AC	AA014051;		
DF	06-JAN-1992	(first entry)	
XX	Remnin gene.		
DE			
XX	Prorennin; alpha-SI-casein gene; insulin-like growth factor I; IGF-I;		
KW	mammary gland; ss.		
XX			
OS	Synthetic.		
PN	EPA51823-A.		
XX			
PD	16-OCT-1991.		
XX			
PE	10-APR-1991; 91EP-0105702.		
XX			
PR	19-APR-1990; 90DE-4012526.		
XX			
PR	11-APR-1990; 90DE-4011751.		
XX			

[illegible]

KW microbial vector; ss.
XX Bos taurus.
XX Key Location/Qualifiers
FT CDS 1..1140
FT /*tag= a
XX BE897201-A.
XX PD 03-NOV-1983.
XX PE 30-JUN-1983; 83BE-0017731.
XX PR 13-APR-1983; 83US-0484539.
XX PR 01-JUL-1982; 82US-0394433.
XX (GENE-) GENEX CORP.
XX WPI; 1983-820813/47.
DR P-PSDB; AAP30013.
XX Isolated chymosin or rennin and prochymosin genes - plasmid(s)
PT which replicate in prokaryotic organisms, esp. *Escherichia coli*,
PT and organisms used for chymosin biosynthesis
PS Claim 7; Page 33-36; 43pp; French.
XX The inventors claim isolated chymosin (rennin) and prochymosin genes
CC from calves, and plasmids contg. the genes which are capable of
CC replicating in a prokaryotic organism. The prokaryotic organism is
CC pref. an *Escherichia* species, esp. *E. coli* p Gx 1225 (NRRL B-15061).
CC The microorganisms transformed by the plasmid are also claimed.
XX
SQ Sequence 1289 BP; 310 A; 391 C; 327 G; 260 T; 1 other:
Query Match 75.9%; Score 889.8; DB 4; Length 1289;
Best Local Similarity 87.2%; Pred. No. 1.6e-270;
Matches 975; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
QY 56 tctgttcgttactcaagctgttgatccgcatctctctctcaagaagtaagtttc 115
DB 23 tcttcgtctctccagcgctgagatccagagatccctctgtacaaggaagcttc 82
QY 116 tccgttaagcgctgaaagacatggaactctagaagactcttcgagaacaacagta 175
DB 83 tgaggaagcgctgagagcatggtctcttgaggaactctctcgagaacagcagttc 142
QY 176 gcatacgaagaagtaactcggcttcggtgaagttgctagcgttgccaacttaccactacc 235
DB 143 gcatcagcagcaagtaactcggcttcggtgagcgttgccagcgtygccctcgaccaatacc 202
QY 236 ttgatatgaataactttgggaagatctactcggaaaccccgctcaagaagttcaccttc 295
DB 203 tagataagtaactcttgggaagatctactcggaaaccccgctcaagaagttcaccttc 262
QY 296 tcttgataactgttctctgacttcgttggttccctctactactactcaagaagcaatgcct 355
DB 263 tgtttgacactgtctctctgacttcgttgtaacctcaatctactactcaagaagcaatgcct 322
QY 356 gcaagaaccaccaagaatcgatccgagaagaagtcgtccaacttcagaacttaggcaaac 415
DB 323 gcaaaaaccaccagcgcttcgacccgagaagaagtcgtccaacttcagaacttaggcaaac 382
QY 416 ccttgtctataactcaggtacaggtagcatgcaagaatcttagctatgataacgtca 475
DB 383 ccttgtctataactcaggtacaggtagcatgcaagaatcttagctatgataacgtca 442
QY 476 ctgtctccaacattgtggacattcaacagacagtaggaacttaggaacgaagaacaggtc 535
DB 443 ctgtctccaacattgtggacattcaacagacagtaggaacttaggaacgaagaacaggtc 502

QY 536 atgtcttaactatgcagaatctgagatgcatccttggtagtgacataccactcgtcgt 595
DB 503 acgttcttaactatgcngaattcgacgggacccggggatggtccatccctcgtcgtcct 562
QY 596 cagagtaactcagatccctgtgtttacaaacatgatgaacccgacactagtaagtaagact 655
DB 563 cagagtaactcagatccctgtgtttacaaacatgatgaacgaaggcactgtgcgccaaagcc 622
QY 656 tgttctcggtttacatggaacgaagaatggcagagagacatgctcaagcttgaagctatgt 715
DB 623 tgttctcggtttacatggaacgaagaatggcagagagacatgctcaagcttgaagctatgt 682
QY 716 atcatccttaactacaagaatcctctcaacttgggttcacgtacactgttgagcagtaactg 775
DB 683 accgctccttaactacaagaatcctctcaacttgggttcacgtacactgttgagcagtaactg 742
QY 776 aattcactgtgacagtgatccacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 835
DB 743 agttcactgtgacagtgatccacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 802
QY 836 aagctatcttgatatacgggtacgttccaaagctgttcggaacttagcagcagactttcaaca 895
DB 803 aggcacatctggaacacgggacactccaaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 862
QY 896 ttcaagaagctatgtgagccacacagaacacagtaacgtgagtttgacatagatgacga 955
DB 863 tccagaagcgcatgtggaacacacagaacacagtaacgtgagtttgacatagatgacga 922
QY 956 accttagctatactgactacagttgtcttggatgaacagcgaagaatgtaaccaactgacc 1015
DB 923 accttagctatactgactacagttgtcttggatgaacagcgaagaatgtaaccaactgacc 982
QY 1016 cctcgcgtatacgaacagagatgaaggtctctgacacagtgagatccaaggttagaac 1075
DB 983 cctcgcgtatacgaacagagatgaaggtctctgacacagtgagatccaaggttagaac 1042
QY 1076 attccagaagaatggtcttggagatgtgttcattcgttgatctactacagcgtcttga 1135
DB 1043 attccagaagaatggtcttggagatgtgttcattcgttgatctactacagcgtcttga 1102
QY 1136 gggccaacaacctgtgtggttagctaaagcaatctga 1173
DB 1103 gggccaacaacctgtgtggttagctaaagcaatctga 1140
RESULT 15
AAQ04683
ID AAQ04683 standard; DNA; 1098 BP.
XX AAQ04683;
AC AAQ04683;
XX 05-OCT-1990 (first entry)
DT 05-OCT-1990 (first entry)
XX Sequence encoding calf pro-rennin.
DE Pro-rennin; ds.
XX Bos taurus.
OS Bos taurus.
XX JP02109984-A.
PN JP02109984-A.
XX 23-APR-1990.
PD 23-APR-1990.
XX 01-JAN-1988; 88JP-0302176.
PF 01-JAN-1988; 88JP-0302176.
XX 01-JAN-1988; 88JP-0302176.
PR 01-JAN-1988; 88JP-0302176.
XX (BEPP/) BEPPU T.
PA WPI; 1990-168358/22.
XX P-PSDB; AAR05080.
DR Complex plasmid and microbe - contains calf pro-rennin cDNA.
XX

XX Disclosure; 32; 13bp; Japanese.
 XX
 CC Protein product may be expressed in E.coli expression system from
 CC plasmid pBR322.
 XX
 SQ Sequence 1098 BP; 252 A; 327 C; 300 G; 219 T; 0 other;

Query Match 75.6%; Score 886.8; DB:11; Length 1098;
 Best Local Similarity 88.0%; Pred. No. 1.3e-269;
 Matches 966; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 76 gctagatcacccgcatctctctacaaagtaagctctccgtaagcgctgaagaa 135
 ||||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 1 gctagatcacccgcatctctctacaaagtaagctctccgtaagcgctgaagaa 60
 Qy 136 catggaactctagaagactctctgcaagaacacagatgagcatcagcaagtaactcc 195
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 61 catggaactctgaggaactctctgcaagaacacagatgagcatcagcaagtaactcc 120
 Qy 196 ggcttcggtgaagtgtagtgtagtgtagtgtagtgtagtgtagtgtagtgtagtg 255
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 121 ggcttcggtgaagtgtagtgtagtgtagtgtagtgtagtgtagtgtagtgtag 180
 Qy 256 aagatctaccctcggaaccccgctcaagaagatcaccgtctctctgtaactggttcctct 315
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 181 aagatctaccctcggaaccccgctcaagaagatcaccgtctctctgtaactggttcctct 240
 Qy 316 gaactctggtctccctctatctactgcaagaagcaatgctctgcaagaacacaaagatlc 375
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 241 gaactctggtctccctctatctactgcaagaagcaatgctctgcaagaacacaaagatlc 300
 Qy 376 gatccggaagaagtgctcaccctcccaacttaaggcaaacccctgctctatacctaagtg 435
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 301 gatccggaagaagtgctcaccctcccaacttaaggcaaacccctgctctatacctaagtg 360
 Qy 436 acaagtgacatgacgaagcaatcttaggctatgatacgtctacactgctccacaactgttgac 495
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 361 acaagtgacatgacgaagcaatcttaggctatgatacgtctacactgctccacaactgttgac 420
 Qy 496 attcaacagacagtagaacttagcaaccaagaacacagtgtagtcttcaactatgcagaa 555
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 421 attcaacagacagtagaacttagcaaccaagaacacagtgtagtcttcaactatgcagaa 480
 Qy 556 ttgcattgcatcctctgtagtgtagttagcaccacgctcgtctagagtagtatacctgtg 615
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 481 ttgcattgcatcctctgtagtgtagttagcaccacgctcgtctagagtagtatacctgtg 540
 Qy 616 ttgacaacatgatacgaacgaacatctagctagctcaagaactgtctcgtttacatgac 675
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 541 ttgacaacatgatacgaacgaacatctagctagctcaagaactgtctcgtttacatgac 600
 Qy 676 aggaatgccaagagagacatgctacagcttgaagctatltgacatcctactacacagaa 735
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 601 aggaatgccaagagagacatgctacagcttgaagctatltgacatcctactacacagaa 660
 Qy 736 tctcttcaactggtctcagctacactgctcagcagtaactggaatcactgtggaagtgctc 795
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 661 tctcttcaactggtctcagctacactgctcagcagtaactggaatcactgtggaagtgctc 720
 Qy 796 acaatcagcggt 855
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 721 acaatcagcggt 780
 Qy 856 acgttcaagctgt 915
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 781 acgttcaagctgt 840
 Qy 916 acacagaacacagtagcgt 975
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 841 acacagaacacagtagcgt 900

Qy 976 gtgtcttggatcaacgcaagatgtagccactgaacccctccgctatacaccagccag 1035
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 901 gtgtcttggatcaacgcaagatgtagccactgaacccctccgctatacaccagccag 960
 Qy 1036 gatcaagggtctcgcaccagtgtagtccagagtgagaacattccacagaatgtgacttg 1095
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 961 gatcaagggtctcgcaccagtgtagtccagagtgagaacattccacagaatgtgacttg 1020
 Qy 1096 ggaagatgttcatctcgtgagtagtactacagcgtcttcttgcaaggccaaacactggtgg 1155
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 1021 ggaagatgttcatctcgtgagtagtactacagcgtcttcttgcaaggccaaacactggtgg 1080
 Qy 1156 ctagctaaagcaatctga 1173
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 1081 ctgagcaaacacatctga 1098

Search completed: July 31, 2002, 21:11:57
 Job time: 25748 sec

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 11:56:19 ; Search time 5848.1 Seconds
(without alignments)
2707.192 Million cell updates/sec

Title: US-09-643-755B-1
Perfect score: 1173
Sequence: 1 atgaacttcttaagctctt.....ggctagctaaagcaatctga 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estln:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrc:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	387.2	33.0	472	10	BG938320
2	358	30.5	631	9	AI892188
3	354.8	30.2	430	10	BG937697
4	337.2	28.7	727	9	AI326975
5	323.8	27.6	399	10	BG937723
6	300.8	25.6	1347	11	AK004109
7	300.6	25.6	383	10	BG938086
8	262	22.3	479	9	AA028632
9	249.6	21.3	505	9	AI479358
10	246	21.0	546	9	AI324867
11	242.4	20.7	517	9	AI385490
12	236.2	20.1	1385	11	AK008959
13	236	20.1	619	9	AW868692
14	227.4	19.4	1388	11	AK008886
15	226.8	19.3	414	9	AI322423
16	217.8	18.6	636	9	AW868716
17	206.4	17.6	814	10	BI761345

18	205.8	17.5	583	10	BE841742
19	204.4	17.4	910	10	BF299798
20	202.8	17.3	622	10	BF373831
21	201	17.1	716	9	AW012992
22	199.4	17.0	603	10	BE841740
23	199	17.0	702	9	AW867433
24	198.6	16.9	551	10	BE841714
25	196.4	16.7	547	10	BM069200
26	196.4	16.7	548	10	BM053683
27	190.6	16.2	391	9	AA969042
28	190	16.2	433	10	W34201
29	190	16.2	713	10	BF121716
30	187.4	16.0	422	10	W10274
31	185	15.8	666	10	BE841761
32	182.6	15.6	594	9	AW863768
33	182.2	15.5	578	10	BE840800
34	182	15.5	564	9	AW868711
35	181	15.4	736	10	BI821121
36	180.8	15.4	616	10	BE841758
37	180.8	15.4	547	10	BE841657
38	180.4	15.4	576	9	AW868673
39	179.6	15.3	960	10	BI517784
40	179.2	15.3	751	9	AI076336
41	179.2	15.3	815	10	BI818940
42	176.8	15.2	815	10	BI818940
43	176.4	15.0	666	9	AW959943
44	176.2	15.0	847	10	BI762154
45	176	15.0	572	10	BF373812

ALIGNMENTS

RESULT 1

LOCUS BG938320 472 bp mRNA linear EST 11-JUN-2001
DEFINITION IAB015E12 Bovine Abomasum cDNA library Bos taurus cDNA 5', mRNA sequence.

ACCESSION BG938320
VERSION BG938320.1 GI:14337692
KEYWORDS EST.

SOURCE

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos. (bases 1 to 472)

REFERENCE 1 (bases 1 to 472)
Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.

TITLE CDNA's from bovine abomasum tissue

JOURNAL Unpublished (2001)

COMMENT Contact: Dr. Stephen Moore

Beef Genomics Laboratory

Dept of AFNS, University of Alberta

410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada

Tel: 780 492 0169

Fax: 780 492 4265

Email: smoores@afns.ualberta.ca

The sequence best matches gb:BOCHYMOA (bovine chymosin a (ferritin mra)) in main database at high score of 928.0 and E-value of 0.0

PCR Primers

FORWARD: M13 Forward

BACKWARD: M13 Reverse

Seq primer: T3 primer

High quality sequence stop: 472

POLYA-NO. Location/Qualifiers

1..472

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="Bovine Abomasum cDNA Library"

/sex="Two males and one female mixed"

/tissue_type="Gastrointestinal tissue (GIT)"

/cell_type="Epithelial"

DB	541	AFATTTGGCAGTGTTCACAGTGGACAGATCATCATCATGTTGAAGTGTGGCTTGTCAAG	600
OY	828	tgtagtcaagctatcttggataccggtac	857
Db	601	TGGCTGCTCCCTGCCGTTCTGGACACAGGCAC	630
RESULT	3		
LOCUS	BC937697	430 bp	mRNA
DEFINITION	LA0005E01	Bovine Abomasum cDNA Library	Bos taurus cDNA 5', mRNA
ACCESSION	BC937697		
VERSION	BC937697.1	GI:14337069	
KEYWORDS	EST.		
SOURCE	COV.		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.		
AUTHORS	Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.		
TITLE	1 (bases 1 to 430)		
JOURNAL	CDNA's from bovine abomasum tissue		
COMMENT	Unpublished (2001)		
	Contact: Dr. Stephen Moore		
	. Beef Genomics Laboratory		
	Dept of AFNS, University of Alberta		
	410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada		
	Tel: 780 492 0169		
	Fax: 780 492 4265		
	Email: smoores@afns.ualberta.ca		
	The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin		
	mRNA) in main database at high score of 844.0 and E-value of 0.0		
	PCR Primers		
	FORWARD: M13 Forward		
	BACKWARD: M13 Reverse		
	Seq primer: T3 primer		
	High quality sequence stop: 430		
	POLYA-No.		
FEATURES			
source	Location/Qualifiers		
	1..430		
	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
	/clone_lib="Bovine Abomasum cDNA Library"		
	/sex="Two males and one female mixed"		
	/tissue_type="Gastrointestinal tissue (GIT)"		
	/cell_type="Epithelial"		
	/dev_stage="Young adult"		
	/lab_host="XLI-BlueMF"-strain		
	/note="Organ: Abomasum; Vector: Uni-22APXR; Site_1: EcoR		
	I; Site_2: Xho I"		
BASE COUNT	97 a 130 c 119 g 84 t		
ORIGIN			
Query Match	30.2%; Score 354.8; DB 10; Length 430;		
Best Local Similarity	89.1%; Pred. No. 3.5e-90;		
Matches 383; Conservative	0; Mismatches 47; Indels 0; Gaps 0;		
OY	605	cgatcccggttttgcacacatgatgaacggcagcactgtaacttaagctgttctcg	664
Db	1	CGATACCCGCTGTTTACACATGATGACAGGACCTGCTGCCCAAGACCTGCTCGG	60
OY	665	tttatactgacaggaatggcagagagacatctcaagcttggaactatgatccact	724
Db	61	TTTACATGGACAGGANTGGCCAGAGAGATCTCACGCTGGGGCCATCGACCCGTCT	120
OY	725	actacacagatcttctcaactgtgttcagatcactgtgtgcagcagtaactgtgaat	784
Db	121	ACTACACAGAGGGCTCTGCATGCTGGCCCGTGCACAGTGCACGACGATGCTGCA	180
OY	785	tggaacagtgtaacctcaagcgggtgtgtgtgtgtgcatgtgaagtgtgagtgtcaagctatct	844

OY	181	TGGACAGTGTCACATFCA6CGGTGTTGTGTGGCTGTGAAGGTGGCTGTGCAGGCCATCC	240
OY	845	tggataccggtatctgccaaagtgtgcgaacctagcagaacattctcaacattcgaaag	904
Db	241	TGGACAGCGGCGACCTTCCAAAGCTGTGTGGGCCCAGACGCACATCTCCAATCCAGCAG	300
OY	905	ctatcgaggccacaagaaccagttacggttgatlttgcacatgatgttcgacaaccttagct	964
Db	301	CCATTGGAGCCACACAGAACCAGTACGATGAGATTGTACATCCACTGCCAACCTGAGCT	360
OY	965	acatctccactcaagttgtcttttgatcatcaacggaagatgtaccactgtgaccctccgct	1024
Db	361	ACATCCCCCATCTGTGCTCTTTGATGATCATGTGCATAAATGTACCCACTGACCCCCTCGCT	420
OY	1025	ataccagcca	1034
Db	421	ATACCAGCCA	430
RESULT	4		
A1326975/c			
LOCUS	A1326975	727 bp	mRNA linear EST 23-DEC-1998
DEFINITION	mj1e03.x1 Soares mouse p3MNF19.5 Mus musculus cDNA clone		
VERSION	IMGE:48128224 3'	similar to TR:Q28075 Q28075 CHYMOSIN C.	; mRNA sequence.
KEYWORDS	A1326975		
SOURCE	A1326975.1	GI:4061404	EST.
ORGANISM	house mouse.		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	1 (bases 1 to 727)		
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubugue,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,B., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The WashU-HMI Mouse EST Project		
JOURNAL	Contact: Marra M/Mouse EST Project		
COMMENT	Unpublished (1996)		
	WashU-HMI Mouse EST Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: mouseest@watson.wustl.edu		
	This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.		
	MGI:293268		
	This clone was previously sequenced on the 5' end only, this new data is from the 3' end		
	Possible reversed clone; similarity on wrong strand		
	High quality sequence stop: 446.		
FEATURES	Location/Qualifiers		
source	1..727		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:482524"		
	/dev_stage="19.5 dpc total fetus"		
	/clone_lib="Soares mouse p3MNF19.5"		
	/note="Vector: DH10B (ampicillin resistant)"		
	/lab_host="DH10B (ampicillin resistant)"		
	/polymerase="PfuI"		
	/primer="Site 1: Not I - oligo(dn) primer 15'		
	"TGTACCAATCTGAAGTGGGCGCGCGCATTTTTTTTTTTTTTTTTTTT 3',		
	double-stranded cDNA was size selected, ligated to Eco RI		
	adapters (Pharmacia), digested with Not I and cloned into		
	(Pharmacia). Library went through one round of		
	normalization to a Cot = 5. Library constructed by Bento		
	Soares and M.Falima Bonaldo. RNA was kindly provided by		
	Dr. Minoru Ku (Wayne State University)."		

BASE COUNT 178 a 180 c 195 g 174 t
ORIGIN

Query Match 28.7%; Score 337.2; DB 9; Length 727;
Best Local Similarity 71.4%; Pred. No. 4.8e-85;
Matches 474; Conservative 0; Mismatches 183; Indels 7; Gaps 2;

```
QY 511 ggaattgacaccagaaccagacgagatgcttcacatcagatcagaatcagatcagatcct 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 727 GGCTGTGATGATACAGAAAGGGGGAGACATCTTAACCTACTCTCCTCATTTGATGGACCTCG 668

QY 571 ggtatgacataccatcgcgtcgcgtcagagatcgcatacctgtgttgacaacatgtg 630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 667 GGCTGTGACATCTTACTTTTGCCTCCAAATACATGATACCATATTTGACAAACATGATG 608

QY 631 aaccgacacccatgtagctcgaagctgtctcgtttacatcagagacgaatgagcagag 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 607 AACAGGACCTGTGGGCCCAAGACCTGTTCCGTTTACATGACGAGAAATGACGAGGGG 548

QY 691 agcatgtcacaagcttgagatcattgacatcctactacacagagatcctcactgagtt 750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 547 AGCATGCTCACACACTGGGGCCATGATGATGCTTACTTCATAGGCTCATGCTGACCTGGTG 488

QY 751 ccactacactgtgcagcagatcgtgcaatcactgtgacagatgtcaacatcagcggtg 810
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 487 CCTGTGACAGTACAGGAGATATTGGCATGTTACAGTGCACAGAGATCATCATGATGTGAA 428

QY 811 gttgttgatgtgaagtgagatgtgcaagctatcttgatagcaggtacgtcgaagctgtc 870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 GTGTGCTTGTCAAGGTGGCTGCTCCCTGCTGACACAGGACGACGCTGTTGACG 368

QY 871 ggaacct-aggcagcagacatcctcaacatcagaagaagatcagagcacaacagaatga 929
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 367 GGGCTGTGATTAAGACATCTCCAAATATTACAGAACTATTGGACGCTGACAGGCCATAA 308

QY 930 cggatgatttgacatagattgcgacaaccttagctacatgctcactagattgtcttgagat 989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 307 TGACCACTTTTGACATGACATGCTGTGGAGGCTGTGACATGATGCCACCGTTGTGTGAGAT 248

QY 990 caagcggaagatgtacacacacagacccctcgcgctacacacagcagagatcaagagttctg 1049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 247 CCATGTGTGAGGAGTCCACCTGACCCCTTATACCAACGAGGTCCAGGCTTCTG 188

QY 1050 caccagtgtatccagagatggaacacatccagaagaatccttgaggatgtgtcatt 1109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 187 CTCGCTGTGGCTTC-----AAGCAGGGGCTCCACATGTGGATCCTTGGGGATGTTTCA 134

QY 1110 tctgtagtactacagcgctcttgacagagggcacaacactcgtgttgactaagcaaat 1169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 TCGGAGATTTCTACACTGTGTTTGACAGAGCCAAACATCGTGTGGGCTGGCCAAAGGCTAT 74

QY 1170 ctga 1173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 73 CTGA 70
```

RESULT 5
LOCUS BG937723 399 bp mRNA linear EST 11-JUN-2001
DEFINITION IAB005D06 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
sequence.
ACCESSION BG937723
VERSION BG937723.1 GI:14337095
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus

REFERENCE Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
AUTHORS CDNA's from bovine abomasum tissue
TITLE

JOURNAL Unpublished (2001)
COMMENT Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@atns.ualberta.ca
The sequence best matches gb:BP019786 (Bos primigenius prochimysin
mRNA, complete cds) in main database at high score of 767.0 and
E-value of 0.0
PCR primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 399
POLYA-No.

FEATURES
Source Location/Qualifiers
1..399
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bovine Abomasum cDNA Library"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="X11-BlueMRF-strain"
/note="Organ: Abomasum; Vector: Uni-22APXR; Site:1: Ecor
I; Site 2: Xho I"

BASE COUNT 89 a 121 c 112 g 77 t
ORIGIN

Query Match 27.6%; Score 323.8; DB 10; Length 399;
Best Local Similarity 88.2%; Pred. No. 2.3e-81;
Matches 352; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
QY 662 cggtttaacatgacaggaatgcccagagagagatcctacgcttbgagctattgtcat 721
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 CGGTTTACATGACAGGAATGGCAGAGAGCATCTCAGCTGGGGCCATCGACCCGT 60

QY 722 cctactacacagatcctctcacttggttcagatcagctgtgcagctactgacatca 781
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 CCTACTACACAGGTCCTCTGCACTGGGTGCCGTACAGTGCAGCAGTACGTGCA 120

QY 782 ctgtgacagtgatcaacatcagcgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 841
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 CTGTGGACAGTGTCAACATCAGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180

QY 842 tcttgatacggatcgttcacagcgtgttcgagccttagacagacattcctcaacattcagc 901
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 TCCTGTGACACGGGACCTTCAGAGCTGTGGGGCCAGCAGCAGCATCTCAACATCCAGC 240

QY 902 aagctattgagcagacacagacacagacagatcagtgatgttgacatagattgcgacaacta 961
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 AGGCCATTGGAGCCACACAGAAACAGTACGATGATGATGATGATGATGATGATGATGATG 300

QY 962 gctacatgctacagttgtcttttgatcaacggaagatgtaaccactgacccctcgcg 1021
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 GGTACATGCCACAGTGTGCTTTGAATCATGCAAAATGTACCACTGACACCCCTCG 360

QY 1022 cctataccagcagatcaaggttcttcgaccagtgat 1060
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 CCTATACCAAGCCAAAGACGAGGCTTCTGTACCAAGTGCT 399
```

RESULT 6
LOCUS AK004109 1347 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus 18 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:1110035E17:PEP5IN08EN F, full insert
sequence.
ACCESSION AK004109

Db	Accession	Source	Length	Score	E-value	Organism
700	AAGGCGACGATGCTGATGCTAGTGGAGTGGATCCCTCTACTACATGAGGAGGCTTAC	GenBank	100	100.0	0.0	Bos taurus
Qy	745 tgggttcacgactcactgtgcagcagagtaactgtgcacatcactgtgcagacgtgtccacac	GenBank	100	100.0	0.0	Bos taurus
Db	760 TGGGTACCAAGTGTCCAAAGCCAGCTACAGGCAATTAGCTGGATAGCATCTCCATGAT	GenBank	100	100.0	0.0	Bos taurus
Qy	805 ggtgtgtgtgtgtcattgtgaaggtgagatcgaacatcattgtgataccgtgacgtcgaag	GenBank	100	100.0	0.0	Bos taurus
Db	820 GGGGAGGTCAATGCTCTGTGATGTGGCTGCCAAGGATATTATGACACACAGGAGCTCTTG	GenBank	100	100.0	0.0	Bos taurus
Qy	865 ctggtcgcagacctagcagcagcattctcacaatcgaacgtatgtgagccacacagaac	GenBank	100	100.0	0.0	Bos taurus
Db	880 CTGACCGGCCCCCAAGCTCCATCGTTAACTATCCAGATCTAATGTGTCGCAAGGCTTCT	GenBank	100	100.0	0.0	Bos taurus
Qy	925 cagtagcgtgtgtgtgtgatatgtgtgcagaacctgtgtacatgacctcagctgtctt	GenBank	100	100.0	0.0	Bos taurus
Db	940 GGTGACGGCGAGTACTTCTCAAGTGTACACACATCAACCTGCTGTGATATGTCTTC	GenBank	100	100.0	0.0	Bos taurus
Qy	985 gagatcaacgcgaagatgttaccactgacccccctccgcctatlacacgacgcagatcaagg	GenBank	100	100.0	0.0	Bos taurus
Db	1000 ACGATCGGCACTGTACTACTACCAAGTGCACGACGATGCTCATCCGAAGATGGTGCA	GenBank	100	100.0	0.0	Bos taurus
Qy	1045 ttctgcaccagttgattccagagtgaac-----cattccagaatggt	GenBank	1089	100.0	0.0	Bos taurus
Db	1060 CACAATTCACAGAGCAACTTTGAGAGGGCATGTAGTACCATCAGACCCTGAGTGTGG	GenBank	1119	100.0	0.0	Bos taurus
Qy	1090 atcttggagagatgtgttcattcgtgtgagctacacagcgtctttagacagccacaacctc	GenBank	1149	100.0	0.0	Bos taurus
Db	1120 GTGCTGGGGGATGTCTTCTCTAGGCTGTATTTCACCCGTGTTGATCGGGCAATAATACAG	GenBank	1179	100.0	0.0	Bos taurus
Qy	1150 gttggctagctaaagcaatctga	GenBank	1173	100.0	0.0	Bos taurus
Db	1180 ATTGCTCGCTCCCTGCTGTCATGA	GenBank	1203	100.0	0.0	Bos taurus

RESULT 7

EG938086
LOCUS
DEFINITION
BOV1011A08 Bovine Abdomasum cDNA Library Bos taurus cDNA 5', mRNA sequence.
EG938086
ACCESSION
VERSION
KEYWORDS
EST.
SOURCE
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 383)
Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
cDNA 5' from bovine Abdomasum tissue
Unpublished (2001)
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/FOR, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0165
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
The sequence best matches gb:BPUI9786 (Bos primigenius prothymosin mRNA, complete cds) in main database at high score of 735.0 and E-value of 0.0

PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 383
POLYA-No.

Location/Qualifiers

1..383
/organism="Bos taurus"
/db_xref="taxon:9913"

BASE COUNT	83 a	111 c	113 g	76 t
ORIGIN				
Query Match	25.6%	Score 300.6;	DB 10;	Length 383;
Best Local Similarity	88.3%;	Pred. No. 9,2e-75;		
Matches 338;	Conservative 0;	Mismatches 44;	Indels 1;	Gaps 1;
QY	599	agtaactcgatacctgtgtgttgacacacatgatacgaacacactagtagtcaagaactgtg	658	
Db	1	AGTACTGATATCCCGTGTGTGACAAATGATGAAACAGACACACTGGTGGCCCAACCTGT	60	
QY	659	tctcggtttacatgacaggaatggccaggaagacatgctcaagcttggagctatgac	718	
Db	61	TCTCGGTTTACATGATGACAGGAATGGCCAGAGAGATGCTCAACGCTGGGGCCATCGACC	120	
QY	719	catctactataacaggatctcttcaactgtgttcaagctacgtgtgacagagactaggaat	778	
Db	121	CGTCTTACTATACACAGGTCCTCGCACTGGGTGCCCGTACATGATGACGACATATGCGACT	180	
QY	779	tcaactgtgacagtgatcaacacacacagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgt	838	
Db	181	TCATCTGTGGACAGTGTACATACATCAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	240	
QY	839	ctatcttggataccggtatgtccaaagctgtgtcggacttagcagcagacattctcaacatc	898	
Db	241	CCATCCTTGAGACACGGGGCACTCCAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	300	
QY	899	agcaagcatttggagacacacacagaaacagctacggtgtgtgtgtgtgtgtgtgtgtgtgt	957	
Db	301	AGCAGGCGCATTTGGAGCCACACAGAACCGATGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	360	
QY	958	cttagctacatgctctacagttgt	980	
Db	361	CTGAGCTATCATGCCACATGTGTGT	383	
RESULT	8			
LOCUS	AA028632	479 bp	mrna	linear
DEFINITION	m14c08.r1 Soares mouse p1NM719.5 Mus musculus cDNA clone IMAGE:463502.5' similar to SW:CHYM_BOVIN P00794 PROCHYMOSIN A AND B			
ACCESSION	AA028632			
VERSION	AA028632.1			
KEYWORDS	EST.			
SOURCE	mouse mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus. 1 (bases 1 to 479)			
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, T., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.			
TITLE	The Mashu-HMMI Mouse EST Project			
JOURNAL	Unpublished (1996)			
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu			

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:277318

Possible reversed clone: similarity on wrong strand
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 444.
Location/Qualifiers

FEATURES

source

1..479
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:463502"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: p773D (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."

BASE COUNT

104 a 127 c 131 g 117 t

ORIGIN

Query Match 22.3%; Score 262; DB 9; Length 479;
Best Local Similarity 71.8%; Pred. No. 1e-63;
Matches 343; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 410 gcaaaccttctctacactgaagtagcatgaagaactctagctatgata 469
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GCAACCCCTCTTTGTCAGATGTGACTGTGATGAGAGGCTTCCGCGCTACGAGA 60
QY 470 ccgtacgtctccacaactgtgacattcaacagacagtagacttagaccacaaga 529
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 CTGTACAGCTCTGTATTTGTATGTGCTCCATCAGACATGTGGCTGTGATACCCAGAAC 120
QY 530 cagtgatgtcttcacactatgcagaattcgatgcatcttgatagcatcaccatgcg 589
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 CGGGGACATCTTACACTACTCTCCATTGATGATGATCGGCGCTTCTTACTACT 180
QY 550 tcggtcagagtagctcgtacctgtgttgacaacatgatgaaccgaacctagtctc 649
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TTGCTCCAAATACATCAATACCATATTTGACATGATGATGAGGCGACCTGGTGCGCC 240
QY 650 aagactgtctcgtttacatgagcagaatgagcagaagacatgctcgaacttgag 709
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AAGACCTTCTTCCGTTTACATGAGCAGGATGAGCAGGAGGACATGCTCACTGAGGG 300
QY 710 ctattgatacactcactacacagaatctctcactggttccagtcactgtgcagcagt 769
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 770 actggaattacacgtgtgacaggtgtcaccatcagcgtgtgtgtgtgtgtgtgtgt 829
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 ATTGGCAATTTACATGAGCAGGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 830 gatgtcaagctatctgtataccgttacgtcgaagctgtgtcgaactgacagcagcat 887
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GCTGCCCTCGCTTGTGACACAGGCACTGCGCTTGTGACGAGGCGCTGTGAGACAT 478

RESULT 9
AI479358 505 bp mRNA linear EST 14-APR-1999
LOCUS AI479358/c
DEFINITION tm27e07.x1 Soares_NFL.T_GBC.S1 Homo sapiens cDNA clone
IMAGE:2157828 3 similar to YR:Q28950 Q28950 PREPROCHYMOSIN
PRECURSOR ; mRNA sequence.
ACCESSION AI479358

VERSION AI479358.1 GI:4372526
EST.
KEYWORDS human.

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 505)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

REFERENCE

1 (bases 1 to 505)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c98pbp-r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 608 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 367.
Location/Qualifiers

FEATURES

High quality sequence stop: 367.

Location/Qualifiers

source

1..505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2157828"
/clone_lib="Soares_NFL.T_GBC.S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
a modified polylinker. Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NH7, and B-cell
NCI-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.
BASE COUNT 106 a 138 c 141 g 119 t 1 others
ORIGIN

Query Match 21.3%; Score 249.6; DB 9; Length 505;
Best Local Similarity 77.1%; Pred. No. 3.7e-60;
Matches 303; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 781 actgtggacagtgatcaccatcagcgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 840
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 505 ACCCTTGNAAGTGTATCATATACGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 446
QY 841 atcttgatgacagtgatcaccatcagcgggtgtgtgtgtgtgtgtgtgtgtgtgtgt 900
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 445 ATCTGTGACACCGGACCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 386
QY 901 caagctatgtgagcacaacacagatcagtgagtgatgtgacatgacagcaccct 960
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 385 CAGGCGCATTTGAGCCATGCGGCGGCGCATGATGATGATGATGATGATGATGATGAT 326
QY 961 agctacatgctacagtgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 325 ACCAGCATTTCCACAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 266
QY 1021 gctctacacagcagatcaagaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 265 GCTTATACAGCCAGGACCAAGGCGCTTTTGTACACAGAGGTTTCCAGGAGTACTATAGTTCC 206
QY 1081 cagaatgtgactgttgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205 CAGCAGTGTGATCTGTGAGGATGTCTTCAATCTGTGAGGATGTCTTGTGACAGGACC 146
QY 1141 aacaacctgt 1173
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 145 AATAACCGT 113

```

RESULT 10
AI324867/c
LOCUS
DEFINITION
AI324867 546 bp mRNA linear EST 23-DEC-1998
m182901.x1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:473065 3' similar to TR:Q28075 Q28075 CHYMOSIN C. ; mRNA
sequence.
ACCESSION
AI324867 GI:4059296
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus.
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 546)
REFERENCE
Marrin, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The Washu-HHMI Mouse EST Project
Unpublished (1996)
CONTACT: Marra M/Mouse EST Project
Washu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:283809
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 451.
FEATURES
Source
Location/Qualifiers
1..546
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:473065"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I - oligo(dt) primer 15',
TGTTCACATCTGAAGTGGAGCGCGCCGATTTTCTTTTCTTTT 3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldi. RNA was kindly provided by
Dr. Minoru Ko (Mayne State University)."
```

```

Query Match 21.0%; Score 246; DB 9; Length 546;
Best Local Similarity 70.9%; Pred. No. 4,1e-59;
Matches 343; Conservative 0; Mismatches 135; Indels 6; Gaps 1;

OY 690 gagacgtcagcgtgagcgtatgatccatctactacacagatctcttcaactgggt 749
|||||
DB 546 GAGCATGCTCACACATGGGGGCATGATCATCTTCAATAGGCTCACTGCATGGGT 487
|||||
OY 750 tccagtcactgtgcagcagctactgcaattcactgtgacagtgtccatcaagcggtt 809
|||||
DB 486 GCCTGTGACAGTACAGGAGGATATTGACAGTTTACAGAGTACATCAATGCTGA 427
|||||
OY 810 gattgttgatctgtagtgagtgatgtcaagctatcttgatctacggtacgtcaagctgt 869
|||||
DB 426 AGTGTGCTGTGTCAAGTGGCTGCCCTCGCTTCTTGACACAGGCAATGCCCTGTTGAC 367
|||||

FEATURES
Source
Location/Qualifiers
1..517
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:482524"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer 15'

OY 870 cgaacctagcagcgacatcttcaacatctcagcagctatiggaagccaacagcagta 929
|||||
DB 366 GGGGCTGTGTGAGACATCTCCATATTCACAGATGATGAGCTGTGCAAGGCCATA 307
|||||
OY 930 cggtagattgacatagatgcgaacaacttagtactacagttgtctttagat 989
|||||
DB 306 TGACCATGTTTACATCCAGCTCTGTGAGGCTGTGACATCATGCCACCGTGTCTTGAGAT 247
|||||
OY 990 caacgcaagatgtaaccactgacccccctcgccctataccagcagatcaagggtctg 1049
|||||
DB 246 CCATGTGTAGGAGATTCACATCCACCTACGCTTATACCAACCAAGGTACAGGCTCTTG 187
|||||
OY 1050 caccagtgatccagagtgtagaacatctcccaaaatggatcttggaagatgtgtcat 1109
|||||
DB 186 CTCACGTGGCTTC-----AAGCAGGCTCCACATGTGGATCTTGCGATGTTTCAT 133
|||||
OY 1110 tctgtagtactacagcgctcttgcagaaggccaacacctgtgtggtactaagaacat 1169
|||||
DB 132 TCGGAGTTATACAGTGTGTTTGACAGAGCCACATGCTGTGGCGTGGCCAGGCTAT 73
|||||
OY 1170 ctga 1173
|||||
DB 72 CTGA 69

RESULT 11
AI385490 517 bp mRNA linear EST 27-JAN-1999
LOCUS
DEFINITION
m181e03.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:482524 5' similar to TR:Q28950 Q28950 PREPROCHYMOSIN
PRECURSOR ; mRNA sequence.
ACCESSION
AI385490 GI:4198953
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 517)
REFERENCE
Marrin, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The Washu-NCI Mouse EST Project 1999
Unpublished (1999)
CONTACT: Marra M/Washu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:293268
This read is a RESEQUENCE of a previously sequenced mouse clone.
This read has been verified (found to hit its original self in the
possible reorientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 442.
Location/Qualifiers
1..517
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:482524"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer 15'
```



```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGI:1895666"
/db_xref="taxon:10090"
/clone="2210410L06"
/sex="male"
/tissue_type="stomach"
/clone_lib="Riken full-length enriched mouse cDNA library"
/dev_stage="adult"
66..1256
/note="data source:SPRR, source key:P20142, evidence:ISS
homolog to GASTRICIN PRECURSOR (EC 3.4.23.3) (PEPSINOGEN
C)
putative"
/codon_start=1
/protein_id="BAB25952.1"
/db_xref="GI:12843350"
/translation="MKMMVALLCLPLEALLRVPPKKKSTRETKKEGVAKOFLK
NKRIDPGKTHGKFGDYSLVPEPMAMDSYGEISIGTPONFLVLPDSSNLMV
SSVYCOSEACTTHRYNPKSSYTYQGQFSLQYGLTGFEGDVLVOSIQVFN
OERGLSENEPNTMYAOFDINGLAVPGISGAYTALOGMEGLASQPLFVYLG
SOOGNGGQIVFGVDENLYTGLTWPVQELWQITIDFLIGNASGSSGCO
GIYDTGSLVMPAOYLINELLOTIGAOGEPGYFVSCDSVSLPTLFVINGVOPFL
SPSSYIIQEBGSCMWGLESLNAESQPLMIGDVPKRSYASSTWATGMALPLLS
RDI"
polyA_signal
1368..1373
/note="putative"
polyA_site
1388
/note="putative"
BASE COUNT      282 a      441 c      337 g      328 t
ORIGIN

```

```

Query Match      19.4% Score 227.4; DB 11; Length 1388;
Best Local Similarity 53.9%; Pred. No. 1.3e-53;
Matches 546; Conservative 0; Mismatches 446; Indels 21; Gaps 3;

QY      76  gctgagatcacccgcatctctctacaaagtaagtcctcgtlaagcgcgtgaagaa 135
      111  GCAGCTTTGATCAGGGTCCCGCCGAGAAATGAAGATGCCGGAGACCATGAGGAA 170
      136  catgactctcagaagactcttcgagaagaacagatgcatcagcagaagactacc 195
      171  CAAAGTGTCTCAAAAGACTTTCGAAAGAACCAAGATGACCCCTGGCAGAAATACAC 230
      196  gctcgtgtagagtcgtcagcgtgcacttaacaa-----ctacctgatatgcaa 246
      231  TTTGGCAAGTTTGGTGACTACAGTACTATAGCCCATGGCCTATATGATGCTTCC 290
      247  tacttggaagaactctacctcgaaacccgcctcaagaagatcacgcttctatgatac 306
      291  TACTATGTTGATCAGCATCGGAGACTCCACCCGAAACTTCGTCCTTTGACACT 350
      307  gtttcctcgtactctggttcctctatctactgcaagagcaatgctcgtaagaacac 366
      351  GGGCTCTCCACACCTGTGGGTCTTCTGTCTACTGCAAGAGAGGCTTGACACACAC 410
      367  caaagatcgatccgagaagatcgtaacacttcagaacttaggcacaaacctgtctata 426
      411  ACCGGCTACAAACCCAGCAATCTCTCCACTTACACTCAAGGGCAGACCTTCTCCCTG 470
      427  cactcaggtacaggtacgatacgaagaacttaggctatgatacgcgtcatcgttccaac 486
      471  CAGTACGGCAGCCGAGCCTTACCGGCTTCGCGTATGACACTGTGAGAGTCCAAAGC 530
      487  attgtgaacattcaacagacagtagacttagcaccacgaacacagtgatgtcttcacc 546
      531  ATCCAGGTCCCTAACCCAGGAGTTGGCTTGAGTAGAATGAGCCTGGACCAAAATTTGTC 590
      547  tatcgaattcgaatgcactcctgatagcatatccatcgctgcgctcagagttactcg 606
      591  TAGCCCAATTATGACGGGATATGGGCTGGCCTACCCGGGCTGTCTTACAGGGGGCC 650

```

```

QY      607  atacctgtgttgacaacatgataaccgacacacttagtcaagactgtcttcggtt 666
      651  ACCACCCCTTCGAGGGCATGTTGGGGAGAGGCGCTGTGTCACAGCCCTTCGTGGTGC 710
      667  tacttgacag-----gaatgacagagagacatgtctaaagcttgagcattgatcca 720
      711  TACTTGGCAGCCAGCAGGGGCTTAACGGCGGGCAGATTTGTTGGGTGGCGTGACGAG 770
      721  tccactacaacagatctctcaactggttgactgaactcgtgcacagtaactgcaatc 780
      771  AACCTGTACACTGGCAGAGCTTACCTGGATTCTCTCAACCCAGAGAGCTTACTGGCAGATC 830
      781  actgtgacagatgacacatcagcagtggtgtgtgttcacg-----tgaagtgatgtc 834
      831  ACCATTTGACGACTTCTTATTTGGCAACCGACCTCTGCTGTGCTCTCTCTGCGCTGC 890
      835  caaactctcttgatcaccgtagctccaagctgtgtgcagctagcagcagatctcaac 894
      891  CAAAGCATTTGACACAGGACGACCTCTGCTGTGCTCAATGCTGCGCACTACCTAATGAA 950
      895  attcaagaagctatgtgagccacacagaacacagtaacagtgatgatacagatgcgac 954
      951  CTCTCGAGACCATAGAGAGCCCGAGAGAGATGAGACGATTTTGTGACCTGCGAC 1010
      955  aacctagctacatgcctacagtgctgtcttgagatacaacggaagatgtaccacgtacc 1014
      1011  AGCGTCAGTACGCTGCTTACCTCCTCATCTTGTGCTCAATGATGATGTCAGATTCCTCC 1070
      1015  cctccgctataccagatcacaagagtgcttcgacacagtgatccagag 1067
      1071  CCTCTCTCATCATCATCAGAGAGAGGCTCTGTCATGCTGAGTGGGTGAGAG 1123

```

```

RESULT 15
AI322423
LOCUS      414 bp      mRNA      linear      EST 23-DEC-1998
DEFINITION m.182b01.y1 Soares mouse p3NMPL9.5 Mus musculus cDNA clone
IMAGE:473065 5' similar to TR:Q28075 Q28075 CHYMOSIN C.; mRNA
sequence.
ACCESSION  AI322423
VERSION     AI322423.1 GI:4056852
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   1 (bases 1 to 414)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
TITLE       JOURNAL
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LML; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:283809
            This read is a RESEQUENCE of a previously sequenced mouse clone
            correct orientation)
            putative full length read
            vector to vector length is 655
            Seq primer: -40RP from Gibco
            High quality sequence stop: 395.
FEATURES
Location/Qualifiers
I..414

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 08:54:58 ; Search time 31.6 Seconds
(without alignments)
1370.847 Million cell updates/sec

Title: US-09-643-755b-2

Perfect score: 2064
Sequence: 1 MNFLKSPFYAFLCFGQYFV.....REYVSVDRAINVGLAKAI 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2064	100.0	390	22	AAU00536
2	1933	93.7	458	10	AA094376
3	1928	93.4	545	20	AA133830
4	1922	93.1	381	5	AA040559
5	1921	93.1	375	5	AA040078
6	1919	93.0	381	5	AA040218
7	1918	92.9	365	10	AA094144
8	1918	92.9	381	4	AA030446
9	1914	92.7	380	3	AA020038
10	1910	92.5	365	11	AA050080
11	1907	92.4	365	4	AA030603

12	1903	92.2	379	4	AA030013
13	1903	92.2	381	4	AA030086
14	1888	91.5	381	13	AA020730
15	1874	90.8	450	10	AA094370
16	1082.5	52.4	388	22	AA027708
17	1031	50.0	925	22	AB053391
18	1008.5	48.9	326	22	AA066589
19	1008.5	48.9	326	22	AA061351
20	995	48.2	339	20	AA032058
21	940.5	45.6	339	22	AB053392
22	869	42.1	388	22	AA047998
23	859	41.6	376	20	AA032035
24	846	41.0	375	20	AA032053
25	846	41.0	391	20	AA032056
26	845.5	41.0	387	20	AA032052
27	831	40.3	165	14	AA042838
28	794.5	38.5	410	22	AA016772
29	792.5	38.4	412	16	AA074207
30	792.5	38.4	412	19	AA071369
31	792.5	38.4	412	20	AA06478
32	792.5	38.4	412	21	AA03685
33	788.5	38.2	376	20	AA032054
34	780.5	37.8	349	20	AA031744
35	773.5	37.5	366	22	AB053394
36	765	37.1	380	20	AA032041
37	764	37.0	377	20	AA032037
38	761	36.9	380	20	AA032043
39	760	36.8	380	20	AA032044
40	758.5	36.7	381	20	AA032046
41	754	36.5	380	20	AA032048
42	746.5	36.2	381	20	AA032051
43	744	36.0	380	20	AA032049
44	744	36.0	380	20	AA032050
45	736	35.7	380	20	AA032045

ALIGNMENTS

RESULT 1	
AAU00536	AAU00536 standard; Protein: 390 AA.
ID	AAU00536; (first entry)
AC	AAU00536;
XX	
DT	14-MAY-2001 (first entry)
XX	
DE	Bovine chymosin polypeptide sequence.
XX	
KW	Chymosin; transcription regulator; terminator sequence; soybean; corn; rape seed; sunflower; cotton; tobacco; alfalfa; wheat; barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice; safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander; squash; jojoba.
KW	
XX	
OS	Bos sp.
XX	
FT	Key
FT	Peptide
FT	Peptide
FT	Peptide
FT	Protein
XX	
XX	Location/Qualifiers
XX	1..26
XX	/note= "Signal peptide"
XX	27..67
XX	/note= "Pro sequence"
XX	68..390
XX	/note= "Mature chymosin"
XX	
XX	MO200114571-A1.
XX	
XX	01-MAR-2001.
XX	
XX	23-AUG-2000; 2000WO-CA00975.
XX	
XX	23-AUG-1999; 99US-0378696.
XX	

Sequence encoded b
Sequence encoded b
Prochymosin (prore
Sequence encoded b
Human full-length
Novel human diagno
Human pepsin. Hom
Pepsin protein. H
Cat pregnancy asso
Novel human diagno
Human gastricsin c
Bovine pregnancy a
Bovine pregnancy a
Bovine pregnancy a
Bovine pregnancy a
Prochymosin N-term
Human gene 1 encod
Human death associ
Death associated p
Human tumour-assoc
Amino acid sequenc
Bovine pregnancy a
Human protease HPR
Novel human diagno
Bovine pregnancy a
Bovine pregnancy a
Bovine pregnancy a
Bovine pregnancy a
Bovine pregnancy a
Bovine pregnancy a

PA (SEMB-) SEMBIOSYS GENETICS INC.
XX
PI Van Rooijen G, Keon RG, Boothe J, Shen Y;
XX
XX WPI: 2001-226621/23.
DR N-PSDB: AAS00569, AAS00570.
XX
XX Producing chymosin in seeds of plants such as rice, flax, rape seed, by
PT transforming plant cell with a nucleic acid encoding chymosin operably
PT linked to transcription regulator and terminator sequences -
XX
PS Example 1: Fig 1: 56pp: English.
XX
XX The sequence represents a bovine chymosin polypeptide. Chymosin can be
CC produced in a plant seed through introduction of a chimeric nucleic acid
CC molecule, comprising a nucleic acid sequence encoding a chymosin
CC polypeptide operatively linked to transcription regulator and terminator
CC sequences, into a plant cell. The sequences are useful for producing
CC plant seeds, in particular seeds of soybean, rape seed, sunflower,
CC cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, Arabidopsis
CC thaliana, potato, flax/linseed, safflower, oil palm, groundnut, Brazil
CC nut, coconut, castor, coriander, squash, jojoba and rice.
XX
SQ Sequence 390 AA:

Query Match 100.0%; Score 2064; DB 22; Length 390;
Best Local Similarity 100.0%; Pred. No. 6, 6e-187;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNETKSPFFVAFILCEGOYEAVNTHAEITRIPLLYKGSLSRKALKEHGLEDFLOKQOYGI 60
Db 1 mnetksfpfyafilcfgyfvavthaaetrlplykgsalrkalehgledfldkqgygi 60

QY 61 SSKTSGGEVAVPLTNTLDQYFGKITLYGPQEFVPLPTSGSSDFWVPSTYCKSNACK 120
Db 61 ssktsggevasvplntnyldsqyfgklygcpqefvplfctgsdflwvpslycksnack 120

QY 121 NHOFPDRKSTFONLGRPLSIHYGTSGMOGILGYDVTVSNIVDIOQTGLSTQEPGDV 180
Db 121 nhofdpkrkstfqnlgkplsibhygtsgmgilgydvtvsnlvdiqvtglstqepgdv 180

QY 181 FTYAEFGILGMATPSLASEYSIVFDNMNRHLVADLFSVYMDRNGQESMLTGLAIDP 240
Db 181 ftyaeifgllgmaypslaseysiivfdnmnrhlvaqdlfsvymdrngqesmltlgaidp 240

QY 241 SYTSGSLHWVPVTVQOYWFQVDSVTISGVVACEGGOALIDGTSTKLVPSSDIINTIO 300
Db 241 sytsgslhwvpvtvqywfqvdsvtlsgvvaceggcqaalldgtstklvpsddllniq 300

QY 301 QALGATONQYGEFDIDCDNLSYMPVVEIENGKMPPLTPSAVTSODQFCTSGFQSENHS 360
Db 301 qalga tqygefddidcdnlsympvvefengkmppltpsa vtsodqfctsgfqsenshs 360

QY 361 QKWLIGDVFIREYISVFDNRANLVGLAKAI 390
Db 361 qkwlilgdvfireyysvfdnrannlvglakai 390

RESULT 2
AAP94376 2
ID AAP94376 standard; protein: 458 AA.
XX
AC AAP94376;
XX
DT 15-JUN-1990 (first entry)
XX
XX BamHI/SalI insert of Kluyveromyces plasmid PAB309.
XX
XX Kluyveromyces; PDM100PC; chymosin; tissue plasminogen activator.
XX
XX Kluyveromyces lactis.
XX
OS

PN EP301670-A.
XX
XX 01-JAN-1989.
PD
XX
XX 28-JUL-1988; 88EP-0201632.
PF
XX
XX 28-JUL-1987; 87US-0078539.
PR
XX
XX (KONN) GIST-BROCADES NV.
PA
XX
PI van den Berg JA, van Ooyen AJJ, Rietveld K;
XX
XX WPI: 1989-033565/05.
DR
DR N-PSDB: AAN91188.
XX
XX Kluyveromyces host cells for producing polypeptide(s) -
PT used for highly efficient prodn. of eg chymosin tissue
PT plasminogen activator or human serum albumin.
XX
PS Disclosure; : 56pp: English.
XX
XX BamHI/SalI insert gene product, incorporated into plasmids PAB309 with
CC promoters, terminators and a G418 resistance marker fused to a ADHI
CC promoter from S.cerevisiae.
CC The Kluyveromyces expression systems provide highly efficient secretion
CC and processing of a wide variety of proteins.
CC Sequences identical to those published in EP301669.
XX
SQ Sequence 458 AA:

Query Match 93.7%; Score 1933; DB 10; Length 458;
Best Local Similarity 99.7%; Pred. No. 2, 2e-174;
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 24 HAAETTRIPLYKGSLSRKALKEHGLEDFLOKQOYGISKYSGEVAVPLTNYLDSQY 83
Db 24 haaettripllykgsalrkalehgledfldkqgygisksygfgevasvpltnylasqy 151

QY 84 FGKITLYGPQEFVPLPTSGSSDFWVPSTYCKSNACKNHOFPDRKSTFONLGRPLSIH 143
Db 84 fgkitlygtpqefvplfctgsdflwvpslycksnacknhofdpkrkstfqnlgkplsibh 211

QY 152 FGKILYGTTPQEFVPLPTSGSSDFWVPSTYCKSNACKNHOFPDRKSTFONLGRPLSIH 211
Db 152 fgkilygtppqefvplfctgsdflwvpslycksnacknhofdpkrkstfqnlgkplsibh 211

QY 144 YGTSGMOGILGYDVTVSNIVDIOQTGLSTQEPGDVFTYAEFGILGMATPSLASEYSI 203
Db 144 ygtsgmgilgydvtvsnlvdiqvtglstqepgdvftyaefgllgmaypslaseysi 271

QY 204 PVFDNMNRHLVADLFSVYMDRNGQESMLTGLAIDPSYTYTSGSLHWVPVTVQOYWFQV 263
Db 204 pvfdnmnrhlvaqdlfsvymdrngqesmltlgaidsytytsgslhwvpvtvqywfqvtvd 331

QY 264 SVTISGVVACEGGOALIDGTSTKLVPSSDIINTIOAIGATONQYGEFDIDCDNLSYM 323
Db 264 svtlsgvvaceggcqaalldgtstklvpsddllniqqaigatnqygefddidcdnlsym 391

QY 324 PTVVEIENGKMPPLTPSAVTSODQFCTSGFQSENHSQKWLIGDVFIREYISVFDNRANL 383
Db 324 ptvvefengkmppltpsa vtsodqfctsgfqsenshqkwlilgdvfireyysvfdnrannl 451

QY 392 PTVVEIENGKMPPLTPSAVTSODQFCTSGFQSENHSQKWLIGDVFIREYISVFDNRANL 451
Db 392 ptvvefengkmppltpsa vtsodqfctsgfqsenshqkwlilgdvfireyysvfdnrannl 451

QY 384 VGLAKAI 390
Db 452 vglakai 458

RESULT 3
AAI33830 3
ID AAY33830 standard; Protein: 545 AA.
XX
XX AAY33830;
AC
XX
DT 29-NOV-1999 (first entry)
XX
XX Oleosin-spacer-Met-prochymosin amino acid sequence.
DE

XX	oil-body; lipid body; oleosome; spherosome; separation;
KW	fusion protein; heterologous polypeptide; commercial production;
KW	plasmid; oil-body targeting sequence.
XX	
XX	Synthetic.
OS	Arabidopsis thaliana.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..118
FT	/label= Oleosin targeting sequence
FT	119..175
FT	Protein
FT	/label= Oleosin protein
FT	176..181
FT	cleavage_site
FT	/note="Thrombin cleavage site"
FT	182..483
FT	/label= Met-Prochymosin
XX	
PN	US5948682-A.
XX	
PD	07-SEP-1999.
XX	
PE	25-APR-1997; 97US-0846021.
XX	
PR	25-APR-1997; 97US-0846021.
PR	22-FEB-1991; 91US-0659835.
PR	16-NOV-1993; 93US-0142418.
PR	30-DEC-1994; 94US-0366783.
XX	
PA	(SEMB-) SEMBIOSYS GENETICS INC.
PI	Moloney MM;
DR	WPI: 1999-517960/43.
DR	N-PSDB; AAZ06463.
PT	Expression of a heterologous polypeptide on an oil body protein is
XX	useful for the production of e.g. enzymes, antibodies, hormones
PS	Example 10; Fig 6; 48pp; English.
XX	
CC	This is the amino acid sequence of a HindIII fragment containing the
CC	oleosin-spacer-Met-prochymosin sequence. This HindIII fragment was
CC	joined to a nopaline synthase terminator and cloned into binary vector
CC	PGCM1559. The resulting plasmid was called PSBSOPTNT and introduced
CC	into A.tumefaciens. The resulting bacterial strain was used to transform
CC	B.napus plants.
CC	The DNA which encodes a chimeric fusion protein that consists of the
CC	oil-body targeting sequence, a transcription regulation sequence and the
CC	DNA of the protein of interest can be used to produce antibodies,
CC	glycanases, hormones, proteases, protease inhibitors, seed storage
CC	proteins, thrombin inhibitors, hirudin, interleukins, chymosin, cystatin,
CC	xylanase, carp growth hormone, zein or a collagenase.
CC	The enzyme may be cleaved from the oil body protein or used in
CC	association with the oil body fraction.
CC	Allows production of commercially important proteins on a superior scale
CC	to production by conventional systems. The expressed heterologous
CC	protein can be easily separated from host cell components
CC	due to the use of the oil body as a carrier protein.
XX	
XX	Sequence 545 AA;

Query Match	Similarity	93.4%	Score	1928	DB	20	Length	545	
Best Local	Similarity	97.6%	Pred.	No. 8.3e-174					
Matches	368	Conservative	1	Mismatches	6	Indels	2	Gaps	1

QY	16	GQYFVAVTHA--AEITRIPILYKGSILRKALKEHGLEDFLOKQOYGISISKYSFGFEVASV	73
Db	169	gqhtlvpgrsmaetrlpilykksilrkalkelhlfedtlqkqygisiskysgfgfevasv	228
QY	74	PLTNLDQYRGKTYLGRPROEFTYLLFTDGSDFWVPSLYCSNACKNQRDPKRSSTF	133

Dd	229	pluyldesqygkylgpppefevlfdtsgsdftwpslycksnacknhqrfdprkestf	288
Qy	134	QNLKRPLSIHNGTSGMOGILGDTVTVNSNYVDIOQWGLSTOPEBGVFTFAEFDGILGMA	193
Dd	289	qllgkprlsihygtsgmqllgdydcvtvsnlvdlqqlvglystqdegdvlfaetdglgma	348
Qy	194	YPSLASEYSIVFPPNMNRHLVAODLPSVYMDRNGQESMLTLGADPSYVTGSLHWVPVT	253
Dd	349	ypslaseysivpfddmmnrhlvaqdltsvymdrngesmltlgaidspsyvtgslhwvpyt	408
Qy	254	VOQWQOFVDSVTISGVVYACEGGCOALDPTGTSKLYGSPSDILINTQALGATONQYGF	313
Dd	409	vqgwtqfvdsvtlisgyvvaaceggcgaalldgtusklygssdillnlaqaigaqngqet	468
Qy	314	DIDCDNLSTYMTVVEEINGKMYPLTPSAYTSQDQFCTSGROSEHNSQKMLLDVFTIREY	373
Dd	469	didcdnlstymptvveeingkmypltpsaetsydqdgfcctsgfqsenehseqkllldvftirey	528
Qy	374	YSVPDRANLIGLAKAI 390	
Dd	529	ysvtdrannlvglakai 545	

RESULT	4
ID	AAP40559
XX	AAP40559 standard; Protein; 381 AA.
AC	
XX	AAP40559;
DT	04-FEB-1992 (first entry)
DE	Sequence of a polypeptide displaying milk clotting activity.
KW	Cheese-making; recombinant protein; rennet substitute; milk clot.
XX	
OS	Bos taurus.
XX	
FH	Key
FT	Peptide
FT	Protein
XX	
PN	EPI23928-A.
XX	
PD	07-NOV-1984.
XX	
PF	30-MAR-1984; 84EP-010351.
PR	31-MAR-1983; 83US-0480860.
PA	(CODO-) CODON GENETIC ENG.
PL	Cashion LM, McCaman MT, Rice CW, Sias SR;
DR	WPI; 1984-277277/45.
DR	N-PSDB; AAN40295.
XX	
PT	Recombinant DNA coding for milk clotting polypeptide - which is expressed in transformed bacteria
PS	Claim 10; Fig 2; 39pp; English.
CC	Also claimed is E.coli JM83/PLC7 (ATCC 39325) which is transformed with PLC7 contg. the prorennin derived sequence fused in phase with B-galactosidase. The PLC7 prorennin expression plasmid includes sequences which code for both the pseudorennin and mature rennin cleavage sites between Aas 28-29 and Aas 42-43, respectively.
XX	
SQ	Sequence 381 AA;
Query Match	93.1%; Score 1922; DB 5; Length 381;
Best Local Similarity	98.4%; Pred. No. 1.8e-173;

Matches 364; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 21 AVTHAAETRIPLKKGSLRKALKKEHLLLEDLQKQOYGISSKYSGFGEVASVPLTWYLD 80
 ::
 Db 12 alsqgaeltirplykkslrkalkkehllledflkqgqysksygfgevasvpltnyld 71

QY 81 SOYFGKITLTPPOEFVLTFTDGTSSDFWVPSIYCKSNACKNHOFDRKSSTFONLKP 140
 ::
 Db 72 sqyfgkilylgtppqefvlifdgtssdfwvpsiycksnacknhqfdprkssltfnlgkpl 131

QY 141 SIHGTGSMOGLIGDVTYVSNIVDIQOTVGLSTOEPGDVFTVAEFDGILGMAYPSLASE 200
 ::
 Db 132 sihygtgsmgllgydvtvsnivdiqvtglstqepgdvftyaefdgilgmaypslase 191

QY 201 YSIFVFNMMNRHLVAODLESVYMDRNGQESMLTLGALDPSYTGSLHWVTVYQOYWOE 260
 ::
 Db 192 yslpfdmnmnrhlvagdlsfvyndrngesmltlgaldpsytcslhwvvtvqywf 251

QY 261 TVDSVTISGVVAVCEGCGQALIDTGTSLKLVGPSSDILNIQQAIGATONQYGEFDIDCDNL 320
 ::
 Db 252 tvdsvtisgvvavcegcqalldtgtkslvgpsddlniqgaigatqnydefdidcdnl 311

QY 321 SYMPTVFEINGKMYPLTPSAVTSODGFCSTGFQSENHSQKWLIGDVFIREFYSVDPRA 380
 ::
 Db 312 symptvfeingkmypitpsavtsodgfcstgfqsenhsqkwillgdvfireyysvidra 371

QY 381 NNLVGLAKAI 390
 ::::::::::
 Db 372 nnlvglakai 381

RESULT 5
 AAP40078
 ID AAP40078 standard; Protein: 375 AA.

XX AAP40078:

XX 02-FEB-1992 (first entry)

XX Sequence encoded by prochymosin gene.

XX Prochymosin expression vector; E.coli trp operon; chymosin.

XX Bos taurus.

XX EP121775-A.

XX 17-OCT-1984.

XX 07-MAR-1984; 84EP-0102451.

XX 09-MAR-1983; 83JP-0038439.

XX (BEPF/) BEPPU T.

XX Beppu T, Dozumi T, Nishimori K, Shimizu N, Kawaguchi Y;

XX Hidaka M;

XX WPI: 1984-258001/42.

XX N-PSDB; AAN40055.

XX Expression plasmid comprising prochymosin gene and vector -

XX useful for transforming Escherichia coli for prochymosin prodn.

XX PS Disclosure; Fig 1; 59pp; English.

XX The inventors claim the prochymosin gene comprising a nucleotide
 CC sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or
 CC (b) the 5th codon (CGG) to the 365th codon (ATC); and recombinant
 CC plasmids harboured by Escherichia coli strains deposited as FERB BP-
 CC 262, -263 and -264. Any portion of the nucleotide sequence as
 CC described in AAN40055 can be used. Also claimed is a vector derived
 CC from plasmid pBR322. Typically plasmid pCR501 is obtd. from pOCT 2.

CC The transcriptional direction of pOCT 3 is opposite to that of
 CC pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter
 CC clockwise in pOCT 3.

XX Sequence 375 AA;

Query Match 93.1%; Score 1921; DB 5; Length 375;
 Best Local Similarity 98.4%; Pred. No. 2.2e-173;
 Matches 364; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 21 AVTHAAETRIPLKKGSLRKALKKEHLLLEDLQKQOYGISSKYSGFGEVASVPLTWYLD 80
 ::
 Db 6 alsqgaeltirplykkslrkalkkehllledflkqgqysksygfgevasvpltnyld 65

QY 81 SOYFGKITLTPPOEFVLTFTDGTSSDFWVPSIYCKSNACKNHOFDRKSSTFONLKP 140
 ::
 Db 66 sqyfgkilylgtppqefvlifdgtssdfwvpsiycksnacknhqfdprkssltfnlgkpl 125

QY 141 SIHGTGSMOGLIGDVTYVSNIVDIQOTVGLSTOEPGDVFTVAEFDGILGMAYPSLASE 200
 ::
 Db 126 sihygtgsmgllgydvtvsnivdiqvtglstqepgdvftyaefdgilgmaypslase 185

QY 201 YSIFVFNMMNRHLVAODLESVYMDRNGQESMLTLGALDPSYTGSLHWVTVYQOYWOE 260
 ::
 Db 186 yslpfdmnmnrhlvagdlsfvyndrngesmltlgaldpsytcslhwvvtvqywf 245

QY 261 TVDSVTISGVVAVCEGCGQALIDTGTSLKLVGPSSDILNIQQAIGATONQYGEFDIDCDNL 320
 ::
 Db 246 tvdsvtisgvvavcegcqalldtgtkslvgpsddlniqgaigatqnydefdidcdnl 305

QY 321 SYMPTVFEINGKMYPLTPSAVTSODGFCSTGFQSENHSQKWLIGDVFIREFYSVDPRA 380
 ::
 Db 306 symptvfeingkmypitpsavtsodgfcstgfqsenhsqkwillgdvfireyysvidra 365

QY 381 NNLVGLAKAI 390
 ::::::::::
 Db 366 nnlvglakti 375

RESULT 6
 AAP40218
 ID AAP40218 standard; Protein: 381 AA.

XX AAP40218:

XX 25-JAN-1992 (first entry).

XX Sequence of rennin encoded by recombinant CGF4.

XX Yeast expression vector; GAL1 promoter; Saccharomyces cerevisiae.

XX Bos taurus.

XX GB2137208-A.

XX 03-OCT-1984.

XX 28-FEB-1984; 84GB-0405129.

XX 28-FEB-1983; 83US-0470911.

XX (COLB) COLLABORATIVE RES INC.

XX Botstein D, Davis RW, Fink GR, Taunton-Rigby A, Knowlton RG;

XX Mao JT, Moir DT, Golf CG;

XX WPI: 1984-245517/40.

XX N-PSDB; AAN40180.

XX DNA segment contg. GAL1 promoter linked to gene - useful for

XX direction of expression of the gene in yeast cell

PS Example: Table 4, Page 21-23; 35pp; English.

CC The inventors claim a DNA segment contg. GAL1 promoter linked to
CC gene - useful for direction of expression of the gene in yeast cell.
CC The recombinant material carrying a GAL1 promoter of the yeast
CC galactokinase gene may be used in expressing a desired protein, esp.
CC bovine growth hormone, interferon, proteinin or preproteinin, in the
CC yeast cell. Strains of Saccharomyces cerevisiae producing the
CC polypeptides are produced. Yeast strains deposited as ATCC 20643,
CC 20661, 20662 and 20663, strain designations CGY 196, 457, 461 and
CC 528, resp. are new.

XX Sequence 381 AA:

Query Match 93.0%; Score 1919; DB 5; Length 381;
Best Local Similarity 98.4%; Pred. No. 3.5e-173;

Matches 364; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 21 AVTHAAETTRPLPLKGLSRKALKEHGLEDEFLQKQYGISSKYSGFGEVASVPLTNVYD 80
DB 12 alsqgaettriplpykyslrkalkkehglledflqkqygisksygfgevasvpltnyld 71
QY 81 SOYFKITLGPPOEFTVLPDTGSSDFWVPSIYCKSNACKNHQRPDRKSTFQNLGKPL 140
DB 72 sqyfgkilylgppegftvlfdtgsedfwpsiycksnacknhqrfdrkstfqnlgkpl 131
QY 141 STHYGTSGMNOGILGYDTVTSNIVDIOQTGISTQEPGDFVTFYAEFGDILGMAYPSLASE 200
DB 132 slhygtsgmgqllgydvtvsnivdldiqvtglsdqepdvftyaefdgllgmaysprase 191
QY 201 YSIPVDFNMNRHLVAODLFSVYMDRNGQESMLTGAIDPSYTTGSLHWVPVTVQOYMOF 260
DB 192 yslpvdftnmnrhlvaqdlfsvymdrngesmltgaidspsytgslhwvptvqywmof 251
QY 261 TVDSTTIGVYVACGGCOAILDGTSKLVGPSSDILNIQOATGATQNOYGEFDIDCNL 320
DB 252 tvdsttigsyvavacggcaaildgtsklvgpsdillniqatgaigatqnydefdidcdnl 311
QY 321 SYMPVVEINQKMYPLTPSAYTSQDGFCTSGFQSENHSQKWLIGDVFIREYYSVFDRA 380
DB 312 symptvveingkmypitpsaytsdqdgfcstgfsenhsqkwlvgdvfireyysvfdra 371
QY 381 NNLVGLAKAI 390
DB 372 nnlvglakai 381

RESULT 7

AAP94144
ID AAP94144 standard; protein; 365 AA.

XX AAP94144;

XX 07-JUN-1990 (first entry)

XX Prochymosin.

XX Lactic acid bacteria: cheese; Streptococcus cremoris SK112;

XX proteolase; psk112; chymosin; prochymosin.

XX NL8701378-A.

XX 02-JAN-1989.

XX 12-JUN-1987; 87NL-0001378.

XX 12-JUN-1987; 87NL-0001378.

XX (NEZU-) NEDERL INS ZUIVELON.

XX Simons AFM, De Vos WM.

DR WPI, 1989-030097/04.

DR N-PSDB: AAN91157.

PT DNA fragment having region specific for lactic acid bacteria -

PT is contained in plasmid in microorganism used in prodn. of

PT protein and food prodn. eg cheese.

PS Disclosure; 43pp; Dutch.

CC The DNA encoding prochymosin can be cloned into a plasmid (esp. from

CC S. cremoris SK112) and used to produce large amts of the protein by

CC recombinant DNA techniques. This could overcome shortages of prochymosin

CC due to a shortage of calf stomachs and increasing cheese prodn.

CC Prochymosin is also used in prodn. of yoghurt, butter and buttermilk.

CC See also AAP94145 and AAP94146.

XX Sequence 365 AA:

Query Match 92.9%; Score 1918; DB 10; Length 365;
Best Local Similarity 99.7%; Pred. No. 4.1e-173;

Matches 364; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 26 AETTRIPLYKGSILRKALKEHGLEDEFLQKQYGISSKYSGFGEVASVPLTNVYD 85
DB 1 aettriplpykyslrkalkkehglledflqkqygisksygfgevasvpltnyldsqtg 60
QY 86 KIYLGTPPOEFTVLPDTGSSDFWVPSIYCKSNACKNHQRPDRKSTFQNLGKPLSIHNG 145
DB 61 kiylgtppoeftvlfdtgsedfwpsiycksnacknhqrfdrkstfqnlgkplsihng 120
QY 146 TSGMNOGILGYDTVTSNIVDIOQTGISTQEPGDFVTFYAEFGDILGMAYPSLASE 205
DB 121 tsgmgqllgydvtvsnivdldiqvtglsdqepdvftyaefdgllgmayslase 180
QY 206 FDNMNRHLVAODLFSVYMDRNGQESMLTGAIDPSYTTGSLHWVPVTVQOYMOF 265
DB 181 fdnmnrhlvaqdlfsvymdrngesmltgaidspsytgslhwvptvqywmof 240
QY 266 TTSQVAVACGGCOAILDGTSKLVGPSSDILNIQOATGATQNOYGEFDIDCNL 325
DB 241 ttsqvavacggcaaildgtsklvgpsdillniqatgaigatqnydefdidcdnl 300
QY 326 VVEINGKMYPLTPSAYTSQDGFCTSGFQSENHSQKWLIGDVFIREYYSVFDRA 385
DB 301 vveingkmypitpsaytsdqdgfcstgfsenhsqkwlvgdvfireyysvfdra 360
QY 386 LAKAI 390
DB 361 lakai 365

RESULT 8

AAP30446
ID AAP30446 standard; Protein; 381 AA.

XX AAP30446;

XX 03-AUG-1992 (first entry)

XX Sequence encoded by preprochymosin cDNA.

XX Milk-clotting; cheese making; enzyme; zymogen.

XX Cow.

XX Location/Qualifiers

XX 1..16 /label= preprochymosin

XX 17..58 /label= prochymosin

XX 59..381 /label= chymosin

XX GB2100737-A.
PN
XX
XX 06-JAN-1983.
PD
XX
XX 11-JUN-1982; 82GB-0017096.
PF
XX
XX 10-FEB-1982; 82GB-0003907.
PR
XX
XX (CELL-) CELTECH LTD.
PA
XX
PI Carey NH, Harris TUR, Lowe PA, Doel MT, Emtage JS.
XX
XX WPI: 1983-00545K/01.
DR
DR N-PSDB: AAN30209.
XX
XX prodn. or calf stomach chymosin for cheese making - by
PT cultivation of micro-organisms transformed with vector system
PS
XX Claim 41; Fig 4; 26pp; English.
XX
XX The inventors claim a method for the prodn. of calf stomach chymosin
CC for cheese making. Genes and polypeptides for preprochymosin,
CC prochymosin and chymosin are claimed, as are vector systems and a
CC prochymosin primer.
XX
XX Sequence 381 AA;

Query Match 92.9%; Score 1918; DB 4; Length 381;
Best Local Similarity 98.4%; Pred. No. 4.3e-173;
Matches 364; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 21 AVTHAAEITRIPLYGKSLRKALKEHGLEDFLOKQOYGISKYSGFGEVASVPLTNYLD 80
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 12 alsqgaeltirpilygkslrkalkhegllledflqkqygisksygfgevasvpltnyld 71
OY 81 SOYFGKITLGPPOEFFYVLFDTGSSDFWVPSIYCKSNACKNHRDPKRSSTFQNLGKPL 140
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 72 sqyfgkilylgtpgqefvlylftgssdfwvpsiycksnacknhrdpkrkstfqnlgkpl 131
OY 141 SIHYGTGSMOGILGVDYTVSNIVDIOQTGSLSTQEPGDFVFAEFDDGLGMAVPSLASE 200
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 132 slhygtsgmqgllgydltvtsnivaldqlvtglstqepdvfyfaefdgllgmavpslase 191
OY 201 YSIPEFDNMNMNRHLVAODLFSVYMDRNGOESMLTGAIDPSYVTGSLHWPVTVQOYWOQF 260
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 192 ysiptvfdmmnrhlvaqdlfsymdringemsltlgalnpsyytgsllhwpvtvqywgqf 251
OY 261 TVDSVTISGVVACGCGCAIIDLDTGTSKLVGPSSDILNIQQAIGATQNOYGEFDIDCDNL 320
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 252 tvdsvtisgvvacegscgaalldtgsrklvgpsdillnqgaigatqnygefdidcdnl 311
OY 321 SYMPTVVEFINKMPLTPSATYSODOGFCTSGFOSENHSQKMIIGDVFIREYVSVPDRA 380
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 312 symptvvefingkmpltpsatysodogfctsgfsgsenhsqkmiigdvfireyvsvfdra 371
OY 381 NNVLGLAKAI 390
|:|:|||||||||
DB 372 nnlvglakai 381

RESULT 9
AAP20038
ID AAP20038 standard; Protein: 380 AA.
XX
XX AAP20038;
AC
XX 16-DEC-1992 (first entry)
DT
XX
XX Pre-prorennin-A protein sequence.
DE
XX
XX Pre-pro-rennin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;
KW

KW protease; milk-clotting enzyme; ss.
XX
XX Bos taurus.
OS
XX
XX GB2091271-A.
PN
XX
XX 28-JUL-1982.
PD
XX
XX 15-JAN-1982; 82GB-0001120.
PF
XX
XX 01-DEC-1981; 81US-0325481.
PR
XX 16-JAN-1981; 81US-0225717.
XX
XX (COLB) COLLABORATIVE RES INC.
PA
XX
XX Alford BL, Mao J, Moir DT;
PI
XX
XX WPI: 1982-62028E/30 (62028E).
DR
DR P-PSDB: AAP20038.
XX
XX Transformed cells producing rennin and its precursors - contg.
PT appropriate recombinant DNA material
PS
XX Disclosure: Table 1; 39pp; English.
XX
XX DNA sequences either side of the protein sequence
CC can be removed and are not essential to use of the gene in
CC expression. The protein may be expressed in E. coli using
CC plasmid pCGE21. The resulting expressed enzyme is a well
CC known milk-clotting enzyme used in cheese-making.
XX
XX Sequence 380 AA;

Query Match 92.7%; Score 1914; DB 3; Length 380;
Best Local Similarity 98.1%; Pred. No. 1e-172;
Matches 363; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 21 AVTHAAEITRIPLYGKSLRKALKEHGLEDFLOKQOYGISKYSGFGEVASVPLTNYLD 80
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 11 alsqgaeltirpilygkslrkalkhegllledflqkqygisksygfgevasvpltnyld 70
OY 81 SOYFGKITLGPPOEFFYVLFDTGSSDFWVPSIYCKSNACKNHRDPKRSSTFQNLGKPL 140
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 71 sqyfgkilylgtpgqefvlylftgssdfwvpsiycksnacknhrdpkrkstfqnlgkpl 130
OY 141 SIHYGTGSMOGILGVDYTVSNIVDIOQTGSLSTQEPGDFVFAEFDDGLGMAVPSLASE 200
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 131 slhygtsgmqgllgydltvtsnivaldqlvtglstqepdvfyfaefdgllgmavpslase 190
OY 201 YSIPEFDNMNMNRHLVAODLFSVYMDRNGOESMLTGAIDPSYVTGSLHWPVTVQOYWOQF 260
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 191 ysiptvfdmmnrhlvaqdlfsymdringemsltlgalnpsyytgsllhwpvtvqywgqf 250
OY 261 TVDSVTISGVVACGCGCAIIDLDTGTSKLVGPSSDILNIQQAIGATQNOYGEFDIDCDNL 320
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 251 tvdsvtisgvvacegscgaalldtgsrklvgpsdillnqgaigatqnygefdidcdnl 310
OY 321 SYMPTVVEFINKMPLTPSATYSODOGFCTSGFOSENHSQKMIIGDVFIREYVSVPDRA 380
|:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 311 symptvvefingkmpltpsatysodogfctsgfsgsenhsqkmiigdvfireyvsvfdra 370
OY 381 NNVLGLAKAI 390
|:|:|||||||||
DB 371 nnlvglakai 380

RESULT 10
AAR05080
ID AAR05080 standard; protein: 365 AA.
XX
XX AAR05080;
AC
XX

DT 05-OCT-1990 (first entry)
XX
DE Sequence of calf pro-rennin.
XX
KW Pro-rennin; ds.
XX
OS Bos taurus.
XX
PN JP02109984-A.
XX
PD 23-APR-1990.
XX
PF 01-JAN-1988; 88JP-0302176.
XX
PR 01-JAN-1988; 88JP-0302176.
XX
PA (VEPP) VEPPU T.
XX
DR WPI; 1990-168358/22.
XX
DR N-PSDB; AA004683.
XX
PT Complex plasmid and microbe - contains calf pro-rennin cDNA.
XX
PS Disclosure; 32; 13p; Japanese.
XX
CC Protein product may be expressed in E.coli expression system from
XX plasmid pBR322.
XX
SQ Sequence 365 AA;

Query Match	92.58;	Score 1910;	DB 11;	Length 365;
Best Local Similarity	99.28;	Pred. No. 2.3e-172;		
Matches 362;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0;

Qy	26	AEIRIRILYRKKSRLKALKEHGLELDELQKQOQYISKSYGSGFGVNASVPLITNLDSYFC	85
Db	1	aeltfripdykkslrkalkenyllfedfqlqxygirkysgfgvnaapltnldsyf	60
Qy	86	KIYLGPPEQETVFLFDGSSDFWVPSIYCKSNACKNNQRPDKRSYFQNLGKPLSHYG	145
Db	61	klylgtppqetfvlfdgssdfwvpslycksnacknqrtdprktsfqlngkplshyg	120
Qy	146	TGSMOGLIGDYTVTSNIVDIQOTVGLSTPEGDVFTYAEFDGILGNAPYSLASESIPV	205
Db	121	tgsmsgllygtvtevsnlhvdqlqvtgslstqepgqvflyaefdjlglnapyslaseysipv	180
Qy	206	FDNNMNRRLVAQDILFSYIMDRNGOESMLTGADIDPSYITGSLHMVPTVQOYQWQFTVDSV	265
Db	181	fdnnmnrhrlvqadlfsyymdrngesmltgaidspsyitgslhmvpvtyqyqwfctvds	240
Qy	266	TISGVVACBEGGCAIILDGTGRKLVGSSDILINQQAIGATNQNOYGEFDIDCONLSMPT	325
Db	241	tisgvvaaceggcgailldgtgrkslvgpsdlinlqgaigatqnqydeididconlsymp	300
Qy	326	VVFELNGKMPPLTPSAVTSQDQGFCTBGFQSESNHSQKMLIGDVFIREYSVFPFRANLIVG	385
Db	301	vvfelngkmpltpsaytsdqgqfctbsfgqsehsqkmligdvfireysvfdramlvg	360
Qy	386	LAKAI 390	
Db	361	lakti 365	

RESULT	11
AAP30603	
ID	AAP30603 standard; Protein; 365 AA.
XX	
AC	AAP30603;
XX	
DT	14-JUN-1992 (first entry)
XX	
DE	Sequence encoded by proteinin cDNA in PCR 10001

```

XX  Rennin; renin; enzyme; protease.
KW
XX
OS  Bos taurus.
XX
PN  EP73029-A.
XX
PD  02-MAR-1983.
XX
PF  19-AUG-1982; 82EP-0107601.
XX
PR  24-AUG-1981; 81JP-0131631.
XX
PA  (BEEP/) BEEP T.
XX
PI  Beppu T, Uozumi T, Nishimori K;
XX
DR  WPI: 1983-22976K/10.
DR  N-PSDB; AAN30063.
XX
XX  Plasmid contg. calf pro:rennin DNA - and transformed
PT  microorganisms
XX
PS  Example; Page 20-23; 32pp; English.
XX
CC  The inventors claim recombinant plasmids contg. the cDNA of calf
CC  prorennin. Specified plasmids are PCR 10001 and PCR2001 (contg. the
CC  whole sequence plus the lac promoter region). Also new are
CC  microorganisms transformed with the plasmids, esp. E. coli CRI (ATCC
CC  391710) contg. plasmid PCR2001.
XX
SQ  Sequence 365 AA;

```

Query Match	92.4%;	Score 1907;	DB 4;	Length 365;
Best Local Similarity	98.98;	Pred. No. 4.5e-172;		
Matches 361; Conservative	1;	Mismatches 3;	Indels 0;	Gaps 0;

Qy	26	AEITPIPIYKCKSLRKALKKEHGLEEDFLQXQOYGISSSKSGSFGFVNASVPLTNLNLDSQYFG	85
Db	1	aetlttriplxygslrkalkehgllledltlqkxqyirfskysqfgevavpltnlndsqytig	60
Qy	86	KIYLGTPQOEFTVLEFDTSSDFWVPWPIYCKSNACKNNQRPFRKRSYFQNLGKPLSIHYG	145
Db	61	kyltltppqetfvtlfdtsssfwvpslycksnacknqrtpdrktsscfnglnqplshyq	120
Qy	146	TGSMQGLIGYDTVTYVSNIVDIQQTGVLSTOEBGVFTYAEFDGILGNAPSLASESIPV	205
Db	121	tqsmggllygdtvtvsnlvdltqcvglstqpgdvftyaetdglngnapslaseysipv	180
Qy	206	PDNNNNRRLVADDFSVYMDNKGCESMULTGADIDPSTYTGSLHWPVTYQOYQOFTYDSV	255
Db	181	fdnnnnrrllvagdflsvymdnkgcesmlltgaaldpsytgslnhwpvtvyqywfitydsv	240
Qy	266	TISGVVAVCEGGCOALIDTGTSGKLVGSPSDLINTLOALIGATONQYGEFDIDCONLSMPT	325
Db	241	tisgvvavceggcqaalldtgtsklvgpsldlnlntloaligaatnqydefdidconlsymp	300
Qy	326	VFEELNGKMYPLRPSAYTISODQGECTSGOFSEHNHQMLIGDVEIRYYSVFPDRANNTYG	385
Db	301	vfeelngkmyplrpsaytisdqgfcctsgfsehnshqmligdvfeiryysvfdrrannlyg	360
Qy	386	LAKAI 390	
Db	361	lakai 365	

```

RESULT      12
AAP30013
ID    AAP30013  standard; peptide: 379 AA
XX
AC
XX    AAP30013;
XX

```

DT 25-APR-1992 (first entry)
 XX Sequence encoded by veal chymosin gene.
 DE
 XX
 XX
 KW Protolytic enzyme; zymogen; rennin; chymosin; cheese making;
 KW microbial vector.
 OS
 XX Bos taurus.
 PN BE897201-A.
 XX
 XX
 PD 03-NOV-1983.
 XX
 XX 30-JUN-1983; 83BE-0017731.
 PF
 XX 13-APR-1983; 83US-0484539.
 PR 01-JUL-1982; 82US-0394433.
 XX
 XX (GENE-) GENEX CORP.
 PA
 XX WPI: 1983-820813/47.
 DR N-PSDB; AAN30022.
 XX
 PT Isolated chymosin or rennin and prochymosin genes - plasmid(s)
 PT which replicate in prokaryotic organisms, esp. Escherichia coli,
 PT and organisms used for chymosin biosynthesis
 XX
 PS Disclosure; Page 33-36; 43pp; French.
 XX
 CC The inventors claim isolated chymosin (rennin) and prochymosin genes
 CC from calves, and plasmids contg. the genes which are capable of
 CC replicating in a prokaryotic organism. The prokaryotic organism is
 CC pref. an Escherichia species, esp. E. coli p Gx 1225 (NRRL B-15061).
 CC The microorganisms transformed by the plasmid are also claimed.
 XX
 SQ Sequence 379 AA;

Query Match 92.2%; Score 1903; DB 4; Length 379;
 Best Local Similarity 97.8%; Pred. No. 1.1e-171;
 Matches 362; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 21 AVTHAAETRIPLKYGKSLRKALKEHGLEDFLOKQOYGISKSGEGEVASVPLTNYLD 80
 DB 10 alsqgaeltirplykyslrrkalkhehgledfldqkqsgyskysgfdvasvpltnyld 69
 QY 81 SOYFGKITLGPPOEFYVLEFDGSSDFWVPSITCKSNACKNHQRPDKKSTFONLGRPL 140
 DB 70 sqyfgkitlylgtppqefvlyldtgsdltwpslycksnacknhqrdpkrkstlgnldgkpl 129
 QY 141 SIHVGTSMOGILGYDVTYSNIVDIOOTVGLSTOEPGDVFTYAEFDGILGMAVPSLASE 200
 DB 130 slhytgtsmgilgydvtvsnivaigqvtgslstqepgdvftkyefdgilgmaypslase 189
 QY 201 YSIPFEDMMNRHLVAODLFVYMDRNOESMLTLGALDPSTYTGSLHWVPTVQOYWOF 260
 DB 190 ysiplfdmnmrhlvaqdlfsvymdrngesmltlgaldpsytgslhwvptvqyqcf 249
 QY 261 TVDSVTISGVVACGCGOAILDTGTSKLVGPSSDILNIQOALGATNOYGEFIDICNL 320
 DB 250 tvdsvtisgvvacegqcgaildtgtsklvgspsdlinlqdalgtqnygefdidcndl 309
 QY 321 SYMPTVVEINKMKYPLTPSATVSODGFCSTSGFSENHOSKWLIGDVFIREYYSVFDRA 380
 DB 310 symptlvveinkmkyp ltpsaytvsodgfcstsgfsehnshqkwlvgdvfireyysvfdra 369
 QY 381 NNVLGLAKAI 390
 DB 370 nmlvglakai 379

RESULT 13
 AAP30086

ID AAP30086 standard; Protein; 381 AA.
 XX
 AC AAP30086;
 XX
 DT 14-JUN-1992 (first entry)
 XX
 DE Sequence encoded by cDNA sequence corresponding to one of the
 DE allelic forms (B) of bovine preprochymosin.
 XX
 KW Chymosin; enzyme; rennet; cheese.
 XX
 OS Bos taurus.
 XX
 XX
 FT Key Location/Qualifiers
 FT Peptide 1..16
 FT /label= signal
 XX
 PN EP77109-A.
 XX
 XX 20-APR-1983.
 PD
 XX 13-OCT-1982; 82EP-0201272.
 PF
 XX 14-OCT-1981; 81GB-0031004.
 PR
 XX (UNIL) UNILEVER NV.
 PA
 XX
 PI Maat J, Verrips CT, Ledebroer AM, Edens L;
 XX
 XX WPI: 1983-39656K/17.
 DR N-PSDB; AAN30049.
 DR
 XX DNA molecules comprising genes for preprochymosin - used to
 PT transform microorganisms to give strain producing the
 PT prepro-enzyme and its allelic and maturation forms
 XX
 PS Claim 2; Fig 1; 53pp; English.
 XX
 CC Preprochymosin is an intermediate (via prochymosin and
 CC pseudo-chymosin) for the enzyme chymosin, which is the essential milk-
 CC clotting component of rennet and is used in cheese manufacture.
 CC AAN30049 corresp. to mRNA isolated from the fourth stomach of a
 CC preruminant calf (abomasum, Frisian cow).
 XX
 SQ Sequence 381 AA;

Query Match 92.2%; Score 1903; DB 4; Length 381;
 Best Local Similarity 97.6%; Pred. No. 1.1e-171;
 Matches 361; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 21 AVTHAAETRIPLKYGKSLRKALKEHGLEDFLOKQOYGISKSGEGEVASVPLTNYLD 80
 DB 12 alsqgaeltirplykyslrrkalkhehgledfldqkqyskysgfdvasvpltnyld 71
 QY 81 SOYFGKITLGPPOEFYVLEFDGSSDFWVPSITCKSNACKNHQRPDKKSTFONLGRPL 140
 DB 72 sqyfgkitlylgtppqefvlyldtgsdltwpslycksnacknhqrdpkrkstlgnldgkpl 131
 QY 141 SIHVGTSMOGILGYDVTYSNIVDIOOTVGLSTOEPGDVFTYAEFDGILGMAVPSLASE 200
 DB 132 slhytgtsmgilgydvtvsnivaigqvtgslstqepgdvftkyefdgilgmaypslase 191
 QY 201 YSIPFEDMMNRHLVAODLFVYMDRNOESMLTLGALDPSTYTGSLHWVPTVQOYWOF 260
 DB 192 ysiplfdmnmrhlvaqdlfsvymdrngesmltlgaldpsytgslhwvptvqyqcf 251
 QY 261 TVDSVTISGVVACGCGOAILDTGTSKLVGPSSDILNIQOALGATNOYGEFIDICNL 320
 DB 252 tvdsvtisgvvacegqcgaildtgtsklvgspsdlinlqdalgtqnygefdidcndl 311
 QY 321 SYMPTVVEINKMKYPLTPSATVSODGFCSTSGFSENHOSKWLIGDVFIREYYSVFDRA 380
 DB 311 symptlvveinkmkyp ltpsaytvsodgfcstsgfsehnshqkwlvgdvfireyysvfdra 380

Db 312 symplvvfeingkmypitpsaytsqdgqfctsaftsenhsqkwlvgdvfireyvsfdra 371
 QY 381 NNVLGLAKAI 390
 |||||||||
 Db 372 nnlvglakal 381

RESULT 14

AAR20730
 ID AAR20730 standard; Protein: 381 AA.
 AC AAR20730;
 XX 13-MAY-1992 (first entry)
 DT
 DE Prochymosin (prorennin).
 XX
 KM Rennet; zymogen; autocatalytic activation; pseudorennin; milk;
 KW clotting activity; renin A, rennin B.
 XX
 OS Bos taurus.
 PH Key Location/Qualifiers
 FT Protein 1..381
 FT /label= preprorennin
 FT /note= "also known as preprochymosin"
 FT Peptide 1..17
 FT /label= signal_peptide
 FT Cleavage-site 43..44
 FT /note= "autocatalytic cleavage at pH 2"
 FT Cleavage-site 59..60
 FT /note= "autocatalytic cleavage pH 4.7"
 FT MISC-difference 306
 FT /note= "amino acid determining A or B form of rennin"
 FT MISC-difference 218
 FT /note= "Asp in published sequence"
 PN US5082775-A.
 XX 21-JAN-1992.
 PD
 XX 28-OCT-1988; 88US-0263927.
 PF
 XX 28-OCT-1988; 88US-0263927.
 PR 11-MAY-1988; 88US-0263927.
 PR 11-MAY-1984; 84US-0609495.
 PR 12-DEC-1986; 86US-0940199.
 PR 31-MAR-1983; 83US-0480860.
 PR 28-APR-1986; 86US-0856700.
 XX
 PA (BERL-) BERLEX LABS INC.
 XX
 PI McCaman MT, King JF;
 DR MPI: 1992-049149/06.
 DR N-PSDB: AAQ20949.
 XX
 PT Isolating heterologous polypeptide from bacterial inclusion
 PT bodies - by lysing cells, extrn. with nonionic detergent and sepg.
 PT Insoluble polypeptide
 XX
 PS Disclosure; Fig 6; 21pp; English.
 XX
 CC The prorennin (prochymosin) sequence was deduced from the DNA
 CC sequence obtd. by screening a cDNA library with rennin specific
 CC probes derived from the published amino acid sequence of rennin
 CC (Foltmann et al., J. Biol. Chem. 254, 8447-8456 (1979)) (see 223291,2).
 CC The deduced protein sequence was the same as the published sequence
 CC after cleavage of the signal peptide, except for an amino acid change
 CC at residue 218, which may be an artefact of protein sequencing in
 CC the original sequence. The cleavage product of prorennin, rennin is
 CC synthesized in two active forms rennin A, and rennin B. The cleaved
 CC secreted protein undergoes activation in the acidic conditions of the

CC stomach to generate mature rennin by two autocatalytic cleavage events.
 CC Rennin is an active component of rennet which is used to clot milk in
 CC the process of making cheese.
 CC See also AAR22417.
 XX
 SO Sequence 381 AA;

Query Match 91.5%; Score 1888; DB 13; Length 381;
 Best Local Similarity 97.3%; Pred. No. 3e-170;
 Matches 361; Conservative 4; Mismatches 4; Indels 2; Gaps 2;

QY <21 AVTHAAETRIPILYGKSLRKALKEHGLEDFLQKQYGISKSYSGFEVASVPLTNLD 80
 :: |||||
 Db 12 alsqgaeltirpilykgsrlrkalehnglledflqkqgisksygsfgevavsplytnlyd 71
 QY 81 SQYFGKIYLGTPPEFTVLPDTGSSDFWVPSIYCKSNACKNHQRPDKSSTFQNLGKPL 140
 |||||
 Db 72 sqyfgkiylgtppeftvlfdtgsddfwpislycksnacknhqrfdkrsstfqnlgkpl 131
 QY 141 SIHYGTGSMOGLIGDPTVTSNIVDIOQTGVLSTQEPDVTTFYAFDGLIMAY-PSLIAS 199
 |||||
 Db 132 sihygtgsmgllgldvtvsanivdigeitvylstqepdvitlyaeftdglimgaypslias 191
 QY 200 EYSIPVEFDNMNRRHLVAODLFSYVMDRNGQESMLTGAIDPSYTGSLHWVPTVYQOXYWQ 259
 |||||
 Db 192 eysipvfdnmn-rhlvagdlsfymdrngesmltgaidsytsylgslhwvptvtyqywg 250
 QY 260 FTVDVSVTISGVVACESGCGAILDGTSKLVGPSSDIINIOAIGATQNOYGEFDIDCDN 319
 |||||
 Db 251 ftvdsvtisygvvacdgccgailldgtsklvgpsdlinlqgaigtqnyddefdidcdn 310
 QY 320 LSYMPTVYFEINGKMWPLTPSAVTSQDQFCTSGFSQSNHSOKWLGDFTRERYSVFDR 379
 |||||
 Db 311 lsymptvfyengkmypitpsaytsqdgqfctsaftsenhsqkwlvgdvfireyvsfdra 370
 QY 380 ANNVLGLAKAI 390
 |||||||||
 Db 371 annlvglakal 381

RESULT 15
 AAP94370
 ID AAP94370 standard; Protein: 450 AA.
 XX
 AC AAP94370;
 XX
 DT 15-JUN-1990 (first entry)
 DT
 XX
 DE Sequence encoded by BamHI insert from pDM100PC.
 XX
 KW Kluyveromyces; pDM100PC; chymosin; tissue plasminogen activator.
 XX
 PN EP301670-A.
 XX
 PD 01-JAN-1989.
 PD
 XX 28-JUL-1988; 88EP-0201632.
 PF
 XX 28-JUL-1987; 87US-0078539.
 PR
 XX (KONN) GIST-BROCADES NV.
 PA
 XX van den Berg JA, van Ooyen AJJ, Rietveld K;
 PI
 XX MPI: 1989-033565/05.
 DR
 DR N-PSDB: AAN91185.
 XX
 PT Kluyveromyces host cells for producing polypeptide(s) -
 PT used for highly efficient prodn. of eg chymosin tissue
 PT plasminogen activator or human serum albumin.
 XX
 PS Disclosure; : 56pp; English.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 09:16:08 ; Search time 20.9 seconds
(without alignments)
1793.055 million cell updates/sec

Title: US-09-643-755B-2

Perfect score: 2064

Sequence: 1 MNFKSPFVAFICPGQYFV.....REYVSVPDRANLVGLAKAI 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

PIR-71:*
1: pirl:*
2: pirl2:*
3: pirl3:*
4: pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1932	93.6	381	1	chymosin (EC 3.4.2
2	1829	88.6	381	1	chymosin (EC 3.4.2
3	1605	77.8	381	2	chymosin (EC 3.4.2
4	1601	77.6	380	2	chymosin (EC 3.4.2
5	1108.5	53.7	387	2	chymosin (EC 3.4.2
6	1099.5	53.3	385	2	chymosin (EC 3.4.2
7	1094.5	53.0	383	2	chymosin (EC 3.4.2
8	1092.5	52.9	388	1	chymosin (EC 3.4.2
9	1085.5	52.6	388	1	chymosin (EC 3.4.2
10	1084.5	52.5	388	1	chymosin (EC 3.4.2
11	1082.5	52.4	388	1	chymosin (EC 3.4.2
12	1080.5	52.3	388	1	chymosin (EC 3.4.2
13	1078.5	52.3	388	1	chymosin (EC 3.4.2
14	1075.5	52.1	388	1	chymosin (EC 3.4.2
15	1074	52.0	384	2	chymosin (EC 3.4.2
16	1070	51.8	386	1	chymosin (EC 3.4.2
17	1045	50.6	387	2	chymosin (EC 3.4.2
18	1037	50.2	387	2	chymosin (EC 3.4.2
19	1034	50.1	387	2	chymosin (EC 3.4.2
20	1030	49.9	387	2	chymosin (EC 3.4.2
21	1016.5	49.2	382	1	chymosin (EC 3.4.2
22	957.5	46.4	334	2	chymosin (EC 3.4.2
23	942.5	45.1	336	2	chymosin (EC 3.4.2
24	930	44.7	336	2	chymosin (EC 3.4.2
25	923	44.5	336	2	chymosin (EC 3.4.2
26	918	44.2	336	2	chymosin (EC 3.4.2
27	911.5	43.3	331	2	chymosin (EC 3.4.2
28	894	43.2	332	2	chymosin (EC 3.4.2
29	892	43.2	332	1	chymosin (EC 3.4.2

30	891.5	43.2	384	2	A39314	gastricsin (EC 3.4
31	876	42.4	377	1	PEMOCU	gastricsin (EC 3.4
32	869	42.1	388	2	A29937	gastricsin (EC 3.4
33	866	42.0	394	2	B43356	gastricsin (EC 3.4
34	862	41.8	388	2	UC7246	gastricsin (EC 3.4
35	859	41.6	376	2	I45856	gastricsin (EC 3.4
36	843.5	40.9	389	2	JE0371	gastricsin (EC 3.4
37	826.5	40.0	365	2	S66466	gastricsin (EC 3.4
38	818.5	39.7	387	2	I46617	gastricsin (EC 3.4
39	797.5	38.6	398	2	I51185	gastricsin (EC 3.4
40	792.5	38.4	412	1	KHHUD	gastricsin (EC 3.4
41	784.5	38.0	410	1	KHMSD	gastricsin (EC 3.4
42	775	37.5	407	1	KHRTD	gastricsin (EC 3.4
43	747.5	36.2	389	2	I46616	gastricsin (EC 3.4
44	744	36.0	380	2	B41545	gastricsin (EC 3.4
45	728.5	35.3	344	1	KHPGD	gastricsin (EC 3.4

ALIGNMENTS

RESULT 1

CMBO

chymosin (EC 3.4.23.4) precursor - bovine

N:Alternate names: prechymosin; prorennin B

N:Contents: chymosin; prochymosin

C:Species: Bos primigenius taurus (cattle)

C>Date: 24-Apr-1984 #sequence-revision 09-Sep-1994 #text-change 18-Jun-1999

C:Accession: A25631; A93419; A44608; A92259; A91935; A91495; D2434; A00985

R:Hiada, M.; Sasaki, K.; Uozumi, T.; Beppu, T.

Gene 43, 197-203, 1986

A:Title: Cloning and structural analysis of the calf prochymosin gene.

A:Reference number: A25631; MUID:86301873

A:Accession: A25631

A:Molecule type: DNA

A:Residues: 1-16, 'T', 18-381 <HID>

R:Harris, T.J.R.; Lowe, P.A.; Lyons, A.; Thomas, P.G.; Eaton, M.A.W.; Milligan, T.A.; Nucleic Acids Res. 10, 2177-2187, 1982

A:Title: Molecular cloning and nucleotide sequence of cDNA coding for calf prechymosin

A:Reference number: A93419; MUID:82221400

A:Contents: prochymosin B

A:Accession: A93419

A:Molecule type: mRNA

A:Residues: 1-229, 'N', 231-381 <HAR>

R:Nishimori, K.; Kawaguchi, Y.; Hiada, M.; Uozumi, T.; Beppu, T.

J. Biochem. 91, 1085-1088, 1982

A:Title: Nucleotide sequence of calf prorennin cDNA cloned in Escherichia coli.

A:Reference number: A44608; MUID:82189915

A:Accession: A44608

A:Molecule type: mRNA

A:Residues: 7-14, 'X', 16-42, 'L', 44-87, 'N', 89-301, 'D', 303-324, 'I', 326-334, 'G', 336-342, 'A:Cross-references: GB:J00004

A:Note: authors translated the codon TTG for residue 43 as Phe, ACC for residue 82 as S Ala

R:Polmann, B.; Pedersen, V.B.; Kauffman, D.; Wybrandt, G.

J. Biol. Chem. 254, 8447-8456, 1979

A:Title: The primary structure of calf chymosin.

A:Reference number: A92259; MUID:79239460

A:Contents: chymosin B; disulfide bonds

A:Accession: A92259

A:Molecule type: protein

A:Residues: 59-217, 'D', 219-381 <FOL>

R:Polmann, B.; Pedersen, V.B.; Jacobsen, H.; Kauffman, D.; Wybrandt, G.

Proc. Natl. Acad. Sci. U.S.A. 74, 2321-2324, 1977

A:Title: The complete amino acid sequence of prochymosin.

A:Reference number: A44620; MUID:77234648

A:Accession: A44620

A:Molecule type: protein

A:Residues: 17-173, 'T', 175-217, 'D', 219-251, 'Y', 253-381 <FO2>

R:Chang, W.J.; Takahashi, K.

J. Biochem. 76, 467-474, 1974

A:Title: The structure and function of acid proteases. III. Isolation and characteriz

A:Reference number: A91935; MUID:75060332
A:Contents: active site peptides
A:Accession: A91935
A:Molecule type: protein
A:Residues: 92-96;274-278,280 <CH>
A:Note: The authors did not find Lys-279 in their active site peptide
R:Moit, D.; Mao, J.; Schumm, J.W.; Yovis, G.F.; Alford, B.L.; Taunton-Rigby, A.
Gene 19, 127-138, 1982
A:Title: Molecular cloning and characterization of double-stranded cDNA coding for bovin
A:Reference number: A91495; MUID:83054629
A:Contents: prochymosin A
A:Accession: A91495
A:Molecule type: mRNA
A:Residues: 1-301, 'D', '303-381 <MOI>
R:Ichihara, Y.; Sogawa, K.; Takahashi, K.
J. Biochem. 98, 483-492, 1985
A:Title: Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, and
A:Reference number: A22434; MUID:86059312
A:Accession: D22434
A:Molecule type: protein
A:Residues: 1-18, 'XX', '21', 'XX', '24', 'X', '26', 'X', '28', 'X', '30 <ICH>
C:Comment: The sequence of variant B is shown.
C:Comment: Chymosin is synthesized in the mucosa of the abomasum (fourth stomach) of you
C:Comment: Forms A and B are probably allelic variants. Chymosin B is the predominant fo
C:Genetics:
A:Initons: 21/2; 71/3; 111/1; 150/3; 217/2; 255/2; 303/3; 336/3
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; gastric juice; hydrolase; protein digestion; stomach; z
F:1-16/Domain: signal sequence #status experimental <SIG>
F:17-58/Domain: activation peptide #status experimental <APT>
F:59-381/Product: chymosin #status experimental <MPT>
F:92,274/Active site: Asp #status experimental
F:105-110,265-269,308-341/Disulfide bonds: #status experimental

Query Match 93.6%; Score 1932; DB 1; Length 381;
Best Local Similarity 98.9%; Pred. No. 6.3e-145;
Matches 366; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

21 AVTHAETTRIPLYKGRSLRKALKKEHGLEDFLOKQOYGISKYSGEVAVPLTNVLD 80
12 ALSGAEITRIPLYKGRSLRKALKKEHGLEDFLOKQOYGISKYSGEVAVPLTNVLD 71
81 SQYFGKIYLGTPPEQEFVLEFDTGSSDFWPSIYCKSNACKNHQRFDPKRSSTFONLGRPL 140
72 SQYFGKIYLGTPPEQEFVLEFDTGSSDFWPSIYCKSNACKNHQRFDPKRSSTFONLGRPL 131
141 SIHGTGSMOGLIGDYVTVSNIVDIOQTGVLSTQEPGDVTVAFPGIIGMAVPSLASE 200
132 SIHGTGSMOGLIGDYVTVSNIVDIOQTGVLSTQEPGDVTVAFPGIIGMAVPSLASE 191
201 YSIPVFNMMNRHLVAODLFESVYMDRNGQESMLTGAIDPSYTGSLHWPEVTVOQYWOQF 260
192 YSIPVFNMMNRHLVAODLFESVYMDRNGQESMLTGAIDPSYTGSLHWPEVTVOQYWOQF 251
261 TVDSVTTISGVVAVACEGGCOALIDTGTSKLVGPSSDIINIQAIATATONQYGEFIDICDNL 320
252 TVDSVTTISGVVAVACEGGCOALIDTGTSKLVGPSSDIINIQAIATATONQYGEFIDICDNL 311
321 SYMPTVVFELNGKMYPLTPPSAYTSODGFCFSGSGENSEHSHQWILGDVFIREYTSVPDRA 380
312 SYMPTVVFELNGKMYPLTPPSAYTSODGFCFSGSGENSEHSHQWILGDVFIREYTSVPDRA 371
381 NNIVGLAKAI 390
372 NNIVGLAKAI 381

RESULT 2
CMSHB
chymosin (EC 3.4.23.4) B precursor - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 18-Jun-1999

C:Accession: S10996
R:Pungercar, J.; Struckelj, B.; Gabensek, F.; Turk, V.; Kregar, I.
Nucleic Acids Res. 18, 4602, 1990
A:Title: Complete primary structure of lamb preprochymosin deduced from cDNA.
A:Reference number: S10996; MUID:90356410
A:Accession: S10996
A:Molecule type: mRNA
A:Residues: 1-381 <PUN>
A:Cross-references: EMBL:X53037; NID:91373; PIDN:CAA37209.1; PID:91374
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; gastric juice; hydrolase; protein digestion; stomach
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-58/Domain: activation peptide #status predicted <APT>
F:59-381/Product: chymosin B #status predicted <MPT>
F:92,274/Active site: Asp #status predicted
F:105-110,265-269,308-341/Disulfide bonds: #status predicted

Query Match 88.6%; Score 1829; DB 1; Length 381;
Best Local Similarity 93.2%; Pred. No. 8.7e-137;
Matches 345; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

21 AVTHAETTRIPLYKGRSLRKALKKEHGLEDFLOKQOYGISKYSGEVAVPLTNVLD 80
12 ALSGAEITRIPLYKGRSLRKALKKEHGLEDFLOKQOYGVSSSEYSGGEVAVPLTNVLD 71
81 SQYFGKIYLGTPPEQEFVLEFDTGSSDFWPSIYCKSNACKNHQRFDPKRSSTFONLGRPL 140
72 SQYFGKIYLGTPPEQEFVLEFDTGSSDFWPSIYCKSNACKNHQRFDPKRSSTFONLGRPL 131
141 SIHGTGSMOGLIGDYVTVSNIVDIOQTGVLSTQEPGDVTVAFPGIIGMAVPSLASE 200
132 SIHGTGSMOGLIGDYVTVSNIVDIOQTGVLSTQEPGDVTVAFPGIIGMAVPSLASE 191
201 YSIPVFNMMNRHLVAODLFESVYMDRNGQESMLTGAIDPSYTGSLHWPEVTVOQYWOQF 260
192 YSIPVFNMMNRHLVAODLFESVYMDRNGQESMLTGAIDPSYTGSLHWPEVTLOKKWOQF 251
261 TVDSVTTISGVVAVACEGGCOALIDTGTSKLVGPSSDIINIQAIATATONQYGEFIDICDNL 320
252 TVDSVTTISGVVAVACEGGCOALIDTGTSKLVGPSSDIINIQAIATATONQYGEFIDICDNL 311
321 SYMPTVVFELNGKMYPLTPPSAYTSODGFCFSGSGENSEHSHQWILGDVFIREYTSVPDRA 380
312 SYMPTVVFELNGKMYPLTPPSAYTSODGFCFSGSGENSEHSHQWILGDVFIREYTSVPDRA 371
381 NNIVGLAKAI 390
372 NNIVGLAKAI 381

RESULT 3
JC7247
prochymosin - common marmoset
C:Species: Callithrix jacchus (common marmoset)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C:Accession: JC7247
R:Kageyama, T.
J. Biochem. 127, 761-770, 2000
A:Title: New world monkey pepsinogens A and C, and prochymosins. Purification, charac
A:Reference number: JC7245
A:Accession: JC7247
A:Molecule type: mRNA
A:Residues: 1-381 <KAG>
A:Cross-references: DDBJ:AB038386
A:Experimental source: strain NM813
C:Comment: This protein, a zymogen of pepsins, is the major proteolytic enzyme in ver
C:Superfamily: pepsin
C:Keywords: gastric juice; zymogen

Query Match 77.8%; Score 1605; DB 2; Length 381;

Best Local Similarity 81.4%, Pred. No. 4.3e-119;
Matches 301; Conservative 30; Mismatches 39; Indels 0; Gaps 0;

```
OY 21 AVTHAAETIRRLYKGLSRKALKKEHGLEDFLOKQOYGISKYSGEFVASVPLTNLYD 80
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 12 ALSQASGIVRIPLHKGKSLRKALKKEHGLEDFLOKHQHAVSRKHSRREVASSEFLTNLYD 71
OY 81 SOYFGKIYLGTPPOEFTVLEDTGSSDFWVPVSIYCKSNACKNORPDRKSTFQNLKRP 140
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 72 COYFGKIYLGTPPOEFTVLEDTGSSDLWVPVSYCNVACQNNHRRDPKSSFTQNMKDL 131
OY 141 SIHYGTGSKQGLIGDTVTVSNIVDIQOTVGLSTQEPGDVFTYAEFDGILGMAVPSLASE 200
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 132 SIQYGTGSKQGLIGDTVTVSSIVDPHQTVGLSTQEPGDVFTYAEFDGILGMAVPSLASE 191
OY 201 YSIPEFDNMNRHLVAODLFESYMDRNGESMLTGAIDPSYTGSLHWPVTVQOYKMF 260
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 192 YSVPEFDNMNRHLVAODLFESYMSRNEGSMULTGAIIDPSYTGSLHWPVTVQOYKMF 251
OY 261 TVDSVTISGVVACBGGCAIILDTGTSKLVGPSDDLNTLQOAIQATONQYGEFDIDCDML 320
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 252 TVDSVTYVGVVACDGGCAIILDTGTSMLVGPSDDIFNMQAIGATNEGQYGEFDIDCDML 311
OY 321 SYMPVTVVEINCKMTPLPSPATSDQGFCTSGFQSENHSQKWIIGDVFIREYYSVFDRA 380
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 312 SSMPVTVVEINCKMTPLPSPATSDQGFCTSGFQSDSSQOIMIGDVFIREYYSVFDRA 371
OY 381 NNVLGLAKAI 390
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 372 SNVLGLAKAI 381
```

RESULT 4

147176
chymosin (EC 3.4.23.4) precursor - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 20-Jun-2000
C:Accession: I47176
R:Folkmann, B.; Jensen, A.L.; Loenblad, P.; Smidt, E.; Axelssen, N.H.
Comp. Biochem. Physiol. A 68, 9-13, 1981
A:Title: A Developmental Analysis of the Production of Chymosin and Pepsin in Pigs.
A:Reference number: I47176
A:Accession: I47176
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-380 <FOL>
A:Cross-references: EMBL:U14406; NID:g540096; PIDD:AAB08492.1; PID:g540097
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase

Query Match 77.6%; Score 1601; DB 2: Length 380;
Best Local Similarity 80.1%; Pred. No. 8.8e-119;
Matches 297; Conservative 35; Mismatches 39; Indels 0; Gaps 0;

```
OY 20 VAVTHAAETIRPLKGLSRKALKKEHGLEDFLOKQOYGISKYSGEFVASVPLTNLY 79
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10 LALSGSGITRPLKGRKSLRKALKKEHGLEDFLOKPYALSSKYSREFEVASSEFLTNLY 69
OY 80 DSQYFGKIYLGTPPOEFTVLEDTGSSDFWVPVSIYCKSNACKNORPDRKSTFQNLKRP 139
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 70 DTQYFGKIYLGTPPOEFTVLEDTGSSSELMVPSYCKSDACQNNHRRDPKSSFTQNLKRP 129
OY 140 LSIHYGTGSKQGLIGDTVTVSNIVDIQOTVGLSTQEPGDVFTYAEFDGILGMAVPSLASE 199
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 LSIQYGTGSKQGLIGDTVTVSSIVDPHQTVGLSTQEPGDVFTYAEFDGILGMAVPSLASE 189
OY 200 EYSIPEFDNMNRHLVAODLFESYMDRNGESMLTGAIDPSYTGSLHWPVTVQOYKMF 259
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 190 EYTVPEFDNMNRHLVAODLFAYVMSRNDGSMULTGAIIDPSYTGSLHWPVTVQOYKMF 249
OY 260 FTVDVSVTISGVVACBGGCAIILDTGTSKLVGPSDDLNTLQOAIQATONQYGEFDIDCD 319
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 250 FTVDVSVTVINGVACBGGCAIILDTGTSMVLGPSDDLNTLQOAIQATONQYGEFDIDCGS 309
OY 320 LSYMPVTVVEINCKMTPLPSPATSDQGFCTSGFQSENHSQKWIIGDVFIREYYSVFDRA 379
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 LSSMPVTVVEINCKMTPLPSPATSDQGFCTSGFQSDSSQOIMIGDVFIREYYSVFDRA 369
OY 380 NNVLGLAKAI 390
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 370 SNVLGLAKAI 380
```

RESULT 5

JC7245
pepsinogen A - common marmoset
C:Species: Callithrix jacchus (common marmoset)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C:Accession: JC7245
R: Kageyama, T.
J. Biochem. 127, 761-770, 2000
A:Title: New world monkey pepsinogens A and C, and prochymosins. Purification, charac
A:Reference number: JC7245
A:Accession: JC7245
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: DDBJ:AB038384
A:Experimental source: strain NM794
C:Comment: This protein, a zymogen of pepsins, is the major proteolytic enzyme in ver
volution of mammalian orders and families.
C:Superfamily: pepsin
C:Keywords: gastric juice; zymogen

Query Match 53.7%; Score 1108.5; DB 2: Length 387;
Best Local Similarity 58.6%; Pred. No. 7.3e-80;
Matches 215; Conservative 49; Mismatches 94; Indels 9; Gaps 3;

```
OY 30 RIPLYKGLSRKALKKEHGLEDFLOKQOYGISKYSGEF---VASVPLTNLYDSQYFG 85
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 18 KVSILKKSRLKRNLIENHGLEDFLOKNTLIDPASKYTPPOEATMTMINQPLVNTLMEYFG 77
OY 86 KIYLGTPPOEFTVLEDTGSSDFWVPVSIYCKSNACKNORPDRKSTFQNLKPLSIHYG 145
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 TIGIGTPPOEFTVLEDTGSSNLMVPSIYCSPPACQNNHRRDPKSSFTQNLKPLSIHYG 137
OY 146 TSGMOGILGDTVTVSNIVDIQOTVGLSTQEPGDVFTYAEFDGILGMAVPSLASESTIV 205
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 138 TGSMTGILGDTVTVQVGGIADTNQIFGLSETPESFLYSPFQGLGLAVPSISSSGATPV 197
OY 206 PDNMNRHLVAODLFESYMDRNGOE-SMLTGAIDPSYTGSLHWPVTVQOYKMFYVDS 264
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 198 PDNINMODLVSDPLFVSYSSNDQSGSVWFGIDSSYTGSLHWPVTVQOYKMFYVDS 257
OY 265 VITSGVAVACBGGCAIILDTGTSKLVGPSDDLNTLQOAIQATONQYGEFDIDCDNLSTYMP 324
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 258 ITMNGEALVACBGGCAIIVDTGTSLSGPSPILANIOSYIGASENSGEMVAVSCSAISSLP 317
OY 325 TVVEELNGMAYPLTPSPATSDQGFCTSGFQSEN----HSQKWIIGDVFIREYYSVFDRA 380
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 318 DIVFTINGIYVPPASAYTLIDDEGGCTSGFQGNIPFAYGELMILGDAVPIROYFAVFDRA 377
OY 381 NNVLGLA 387
   ||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 378 NNQVGLA 384
```

RESULT 6

JC7575
pepsinogen A - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7575
R: Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001

[illegible]

Query Match	Best Local Similarity	Matches	Conservative	Score	DB 1	Length	DB 2	Score	DB 3	Length	DB 4	Score	DB 5	Length	DB 6	Score	DB 7	Length	DB 8	Score	DB 9	Length	DB 10	Score	DB 11	Length	DB 12	Score	DB 13	Length	DB 14	Score	DB 15	Length	DB 16	Score	DB 17	Length	DB 18	Score	DB 19	Length	DB 20	Score	DB 21	Length	DB 22	Score	DB 23	Length	DB 24	Score	DB 25	Length	DB 26	Score	DB 27	Length	DB 28	Score	DB 29	Length	DB 30	Score	DB 31	Length	DB 32	Score	DB 33	Length	DB 34	Score	DB 35	Length	DB 36	Score	DB 37	Length	DB 38	Score	DB 39	Length	DB 40	Score	DB 41	Length	DB 42	Score	DB 43	Length	DB 44	Score	DB 45	Length	DB 46	Score	DB 47	Length	DB 48	Score	DB 49	Length	DB 50	Score	DB 51	Length	DB 52	Score	DB 53	Length	DB 54	Score	DB 55	Length	DB 56	Score	DB 57	Length	DB 58	Score	DB 59	Length	DB 60	Score	DB 61	Length	DB 62	Score	DB 63	Length	DB 64	Score	DB 65	Length	DB 66	Score	DB 67	Length	DB 68	Score	DB 69	Length	DB 70	Score	DB 71	Length	DB 72	Score	DB 73	Length	DB 74	Score	DB 75	Length	DB 76	Score	DB 77	Length	DB 78	Score	DB 79	Length	DB 80	Score	DB 81	Length	DB 82	Score	DB 83	Length	DB 84	Score	DB 85	Length	DB 86	Score	DB 87	Length	DB 88	Score	DB 89	Length	DB 90	Score	DB 91	Length	DB 92	Score	DB 93	Length	DB 94	Score	DB 95	Length	DB 96	Score	DB 97	Length	DB 98	Score	DB 99	Length	DB 100	Score	DB 101	Length	DB 102	Score	DB 103	Length	DB 104	Score	DB 105	Length	DB 106	Score	DB 107	Length	DB 108	Score	DB 109	Length	DB 110	Score	DB 111	Length	DB 112	Score	DB 113	Length	DB 114	Score	DB 115	Length	DB 116	Score	DB 117	Length	DB 118	Score	DB 119	Length	DB 120	Score	DB 121	Length	DB 122	Score	DB 123	Length	DB 124	Score	DB 125	Length	DB 126	Score	DB 127	Length	DB 128	Score	DB 129	Length	DB 130	Score	DB 131	Length	DB 132	Score	DB 133	Length	DB 134	Score	DB 135	Length	DB 136	Score	DB 137	Length	DB 138	Score	DB 139	Length	DB 140	Score	DB 141	Length	DB 142	Score	DB 143	Length	DB 144	Score	DB 145	Length	DB 146	Score	DB 147	Length	DB 148	Score	DB 149	Length	DB 150	Score	DB 151	Length	DB 152	Score	DB 153	Length	DB 154	Score	DB 155	Length	DB 156	Score	DB 157	Length	DB 158	Score	DB 159	Length	DB 160	Score	DB 161	Length	DB 162	Score	DB 163	Length	DB 164	Score	DB 165	Length	DB 166	Score	DB 167	Length	DB 168	Score	DB 169	Length	DB 170	Score	DB 171	Length	DB 172	Score	DB 173	Length	DB 174	Score	DB 175	Length	DB 176	Score	DB 177	Length	DB 178	Score	DB 179	Length	DB 180	Score	DB 181	Length	DB 182	Score	DB 183	Length	DB 184	Score	DB 185	Length	DB 186	Score	DB 187	Length	DB 188	Score	DB 189	Length	DB 190	Score	DB 191	Length	DB 192	Score	DB 193	Length	DB 194	Score	DB 195	Length	DB 196	Score	DB 197	Length	DB 198	Score	DB 199	Length	DB 200	Score	DB 201	Length	DB 202	Score	DB 203	Length	DB 204	Score	DB 205	Length	DB 206	Score	DB 207	Length	DB 208	Score	DB 209	Length	DB 210	Score	DB 211	Length	DB 212	Score	DB 213	Length	DB 214	Score	DB 215	Length	DB 216	Score	DB 217	Length	DB 218	Score	DB 219	Length	DB 220	Score	DB 221	Length	DB 222	Score	DB 223	Length	DB 224	Score	DB 225	Length	DB 226	Score	DB 227	Length	DB 228	Score	DB 229	Length	DB 230	Score	DB 231
-------------	-----------------------	---------	--------------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------

```
|||||
Db      377 RANNOVGIA 385

RESULT 11
PEHU
Pepsin A (EC 3.4.23.1) 3 precursor [validated] - human
N:Alternate names: pepsinogen A Isozyme 3
C:Species: Homo sapiens (man)
C:Date: 19-Feb-1984 #sequence_revision 19-Feb-1984 #text_change 08-Dec-2000
C:Accession: A00980; PX0023; E22434; E24252; PX0024
R:Sogawa, K.; Fujii-Kuriyama, Y.; Mizukami, Y.; Ichihara, Y.; Takahashi, K.
J. Biol. Chem. 258, 5306-5311, 1983
A:Title: Primary structure of human pepsinogen gene.
A:Reference number: A00980; MUID:83161158
A:Accession: A00980
A:Molecule type: DNA
A:Residues: 1-388 <SOG>
A:Cross-references: GB:J00279
R:Althaus, S.B.P.; Tanji, M.; Kageyama, T.; Takahashi, K.
J. Biochem. 106, 920-927, 1989
A:Title: A comparative study on the NH2-terminal amino acid sequences and some other pro
A:Reference number: PX0023; MUID:90130402
A:Accession: PX0023
A:Molecule type: protein
A:Residues: 16-100 <ATH>
R:Foltmann, B.
FEBS Lett. 241, 69-72, 1988
A:Title: Activation of human pepsinogens.
A:Reference number: S02663; MUID:89065108
A:Accession: S02663
A:Molecule type: protein
A:Residues: 16-68 <FOL>
R:Ichihara, Y.; Sogawa, K.; Takahashi, K.
J. Biochem. 98, 483-492, 1985
A:Title: Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, and
A:Reference number: A22434; MUID:86059312
A:Accession: F22434
A:Molecule type: protein
A:Residues: 1-15, 'XXX', 19-20, 'X', 22, 'XX', 25-26, 'X', 28 <ICH>
R:Evers, M.P.J.; Zelle, B.; Pieper, D.S.; Mager, W.H.; Planta, R.J.; Eriksson, A.W.; Fre
Hum. Genet. 77, 182-187, 1987
A:Title: Molecular cloning of a pair of human pepsinogen A genes which differ by a Glu->
A:Reference number: I54252; MUID:88006181
A:Accession: I54252
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-27, 'F', 29-73 <RES>
A:Cross-references: GB:M27598; NID:g189834; PIDN:AAA36431.1; PID:g189836
C:Genetics:
A:Gene: GDB:PGN3
A:Cross-references: GDB:119482; OMIM:169710
A:Map position: 11q13.1-11q13.5
A:Introns: 19/2; 73/3; 113/1; 152/3; 219/2; 258/2; 306/3; 339/3
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein diges
F:1-15/Domain: signal sequence #status experimental <STIC>
F:16-388/Product: pepsinogen A 3 #status experimental <ZYM>
F:16-62/Domain: activation peptide #status experimental <APT>
F:60-388/Product: pepsin A 3, minor variant #status experimental <MIN>
F:63-388/Product: pepsin A 3 #status experimental <MAT>
F:94,277/Active site: Asp #status predicted
F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match      52.4%; Score 1082.5; DB 1; Length 388;
Best Local Similarity 56.7%; Pred. No. 8.2e-78;
Matches 208; Conservative 52; Mismatches 98; Indels 9; Gaps 3;
```

```
|||||
Db      379 RANNOVGIA 385

RESULT 12
B30142
Pepsin A (EC 3.4.23.1) 4 precursor - human
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1990 #sequence_revision 20-Aug-1994 #text_change 29-Aug-1997
C:Accession: B30142; E22434
R:Evers, M.P.J.; Zelle, B.; Bebelman, J.P.; van Beusechem, V.; Kraakman, L.; Hoffer,
Genomics 4, 232-239, 1989
A:Title: Nucleotide sequence comparison of five human pepsinogen A (PGA) genes: evolu
A:Reference number: A91627; MUID:89233110
A:Accession: B30142
A:Molecule type: DNA
A:Residues: 1-27, 'F', 29-388 <EVE>
A:Note: the authors translated the codon TTC for residue 28 as Leu, GGC for residue 3
R:Ichihara, Y.; Sogawa, K.; Takahashi, K.
J. Biochem. 98, 483-492, 1985
A:Title: Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, a
A:Reference number: A22434; MUID:86059312
A:Accession: E22434
A:Molecule type: protein
A:Residues: 1-15, 'XXX', 19-20, 'X', 22, 'XX', 25-26, 'X', 28 <ICH>
C:Genetics:
A:Gene: GDB:PGA4
A:Cross-references: GDB:119483; OMIM:169720
A:Map position: 11q13-11q13
A:Introns: 19/1; 73/3; 113/1; 152/3; 219/2; 258/2; 306/3; 339/3
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion; zymoge
F:1-15/Domain: signal sequence #status experimental <STIC>
F:16-59/Domain: activation peptide #status experimental <APT>
F:63-388/Product: pepsin A 4 #status predicted <MAT>
F:94,277/Active site: Asp #status predicted
F:107-112,268-272,311-344/Disulfide bonds: #status predicted
F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match      52.3%; Score 1080.5; DB 2; Length 388;
Best Local Similarity 56.6%; Pred. No. 1.2e-77;
Matches 209; Conservative 53; Mismatches 94; Indels 13; Gaps 4;
```

```
30 RPLPKGSLKRAKKEHGLDEDFLOKQGYGSSKSGGEVAVS----PLTNYLDSOYFG 85
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      19 KVLIRKSLRKLTSERGLKDFLKKHNLNPKARKYFPQWEAPTLVDEOPLNLYLMEYFG 78
```

```
86 RYLGTPPOEFLVLPDFTSSDPFVPSIYCKSNACKNHORFDPKSSFTFQNLKPLSIHYG 145
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      79 TIGTGPADFTVLFDFTGSSNLNWPVSVCSSLACTNNHNRFPNEDSSTYQTSSEIYIYG 138
```


F;94_277/Active site: Asp #status predicted
F;107_112,268-272,311-344/Disulfide bonds: #status predicted
F;130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match	52.18;	Score 1075.5;	DB 1;	Length 388;
Best Local Similarity	56.68;	Pred. No. 2.9e-77;		
Matches 209;	Conservative 52;	Mismatches 99;	Indels 9;	Gaps 3.

OY	28	ITRPIYVKGKSLRKALKEHGLEDFLOKOOYGISSKYSGFEVASY----	PLTNYSIDSOY	83
Dd	17	ITYKPLVAKKSLRKLNLSHGLKNDPLKKNHNLNASKYFFPAEAPFTLIDEDQPLENINIDYEV		76
OY	84	FGKIYLGTRPDEFTVLFDTGSSDSEWBPSTYCKSNACKNHORFDPKRKSTFQNLGKPLSH		143
Dd	77	FTGIGIGTPAQNFTYVEDTGSSNLNWPSPSYCYSLACMDHNLFNPODSTRTATSKYTSIT		136
OY	144	YGTSGMOSILGDTVYVSNIVDIOQYVGLSTQIEGDFYTAERPDGLIGMAYPSLASYST		203
Dd	137	YGTSGMTCILGIDYVYKVGGLSDTQNIIGLSETEGEPFLYAPAEFGILGLAYPSISSSGAT		196
OY	204	PVEDMMNRHLYAODLFVSYMDRNGOE-SMUTLCAIDPSSYTGSLHWNPATVOOYMOFTV		262
Dd	197	PVEDMINMORLYSODLSVYLSADDOSSSVYIFEGIDSSYITGSLMWPVSVIEGYMOISV		256
OY	263	DSVTSYGVVAVCEGCGQALIDTGTSTKLVGPSSDILNLINQOAIAGATONOGYGFIDIDCONLST		322
Dd	257	DSITMNGMTIACAGCCQALVDTGTSLLTGPSPILANIQSDIGASSENSDEGMVYSCAIISS		316
OY	323	MFTVVFEEINGKMYPLTPSAVYISQOQFCTSGFOS-----ENHSQKWLIGDVFIREIYISVD		378
Dd	317	LPDIYFETTINGVOYPLRPSAVYILIOSGSGCTSGFQCMDVPTESGELMILIGDVFIRORYFTVD		376
OY	379	RANNLVGLA	387	
Dd	377	RANNNOVGLA	385	

RESULT 15

pepsinogen A - African clawed frog
JC7574

C:Species: Xenopus laevis (African clawed frog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001

C:Accession: JC7574; PC7119

R:Ikutawa, M.; Inokuchi, T.; Kobayashi, K.; Yasunasu, S.
J. Biochem. 129, 147-153, 2001

A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens

A:Reference number: JC7573; MUID:21064922; PMID:11134969

A:Contents: Stomach

h:Accession: JC7574

A:Molecule type: mRNA

A:Residues: 1-384 <IKU>

A:Cross-references: DDBJ:AB045380

A:Accession: PC7119

A:Molecule type: Protein

A:Residues: 16-35;57-76 <IKZ>

C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like

C:Genetics:

A:Gene: Pga

C:Superfamily: pepsin

C:Keywords: stomach; zymogen

Query Match	52.08;	Score 1074;	DB 2;	Length 384;
Best Local Similarity	54.68;	Pred. No. 3.8e-77;		
Matches 200; Conservative	64;	Mismatches 96;	Indels 6;	Gaps 3

[illegible]

Db	76	ISIGHPPEFVIEDTGSANLAMPVSVCSSQACSNHRRFNDQOSTFOQATNTPVSIQGT	135
QY	147	GSMOGLIGDTFVTVYSNVIDQITQVTLSTQEPGDVFTTAEDGILGMAVPSLASEYSLPVF	206
Db	136	GSMSEGLIDPLDVOYGNIDISNMQMGLESESEGSFLYSPDQITGLGFLSPIASAQAPVF	195
QY	207	DNMARRHLVADLFESVYMDRNGOE-SMLTLCAIDPSYTYGSLHWVPVYQOYWOFTVDV	265
Db	196	DNMSOGILPQNLFSVYLSIDSGTGSYVLFEGVDNSYSGSLMNVPLTAETWQITLDSV	255
QY	266	TTSVYVACBEGGCAILDTGTSTKLVGSPSLLINQAMGNOMNYGFEFIDCDLSTVPT	325
Db	256	STNGOVIAVACOSCAQVATVTGISTLMGTPSTPIANIONTIGASODBSNGOVINCINNI	315
QY	326	VFEELINGAMVPLTPSPATYSOOQGFCTSGFQS---ENHSOKMILGDVFIREYVSVPDRAN	381
Db	316	IYFTLINGVQYPLSPSAVYRQNOQCGSSGFQAMNLPJNSGDMILIGDVFIRQYFVFPDRAN	375
QY	382	NLYGLA 387	
Db	376	NVYALA 381	

Search completed: July 31, 2002, 09:21:41
Job time: 333 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 09:21:18 : Search time 13.41 Seconds

(without alignments)
1126.072 Million cell updates/sec

Title: US-09-643-755B-2

Perfect score: 2064

Sequence: 1 MNFLKSPFPFAFLCFCFGYFY.....REYYSVDRANNLVGLAKAI 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1928	93.4	381	1	CHYM_BOVIN
2	1829	88.6	381	1	CHYM_SHEEP
3	1605	77.8	381	1	CHYM_CALJA
4	1108.5	53.7	387	1	PEPA_CALJA
5	1094.5	53.0	383	1	PEPE_CHICK
6	1093.5	52.9	388	1	PEP2_MACRU
7	1085.5	52.6	388	1	PEPA_MACRU
8	1084.5	52.5	388	1	PEP1_MACRU
9	1082.5	52.4	388	1	PEPA_HUMAN
10	1075.5	52.1	388	1	PEP4_MACRU
11	1070	51.8	386	1	PEPA_PIG
12	1045	50.6	387	1	PEP1_RABIT
13	1037	50.2	387	1	PEP4_RABIT
14	1034	50.1	387	1	PEP3_RABIT
15	1029	49.9	387	1	PEP2_RABIT
16	1012.5	49.1	367	1	PEPA_CHICK
17	997	48.3	388	1	PAG_HORSE
18	946	45.8	388	1	PEP1_RABIT
19	930	45.1	396	1	CATE_HUMAN
20	923	44.7	396	1	CATE_RAT
21	918	44.5	398	1	CATE_MOUSE
22	913	44.2	397	1	CATE_MOUSE
23	911.5	44.2	391	1	CATE_MOUSE
24	892	43.2	392	1	PEPC_CAVPO
25	876	42.4	377	1	PEPC_RAT
26	869.5	42.1	377	1	PEPC_MACRU
27	869	42.1	388	1	PEPC_HUMAN
28	866	42.0	394	1	PEPC_CAVPO
29	862	41.8	388	1	PEPC_CALJA
30	859	41.6	376	1	PAG2_BOVIN
31	807	39.1	420	1	PAG2_PIG
32	797.5	38.6	412	1	CATD_CHICK
33	792.5	38.4	412	1	CATD_HUMAN

ALIGNMENTS

34	784.5	38.0	410	1	CATD_MOUSE
35	775	37.5	407	1	CATD_RAT
36	772.5	37.4	390	1	CATD_BOVIN
37	747.5	36.2	389	1	PAG1_PIG
38	744	36.0	380	1	PAG1_BOVIN
39	728	35.3	382	1	PAG1_SHEEP
40	718	34.8	345	1	CATD_PIG
41	699	33.9	387	1	ASPP_AEDAE
42	691.5	33.5	365	1	CATD_SHEEP
43	675.5	32.7	402	1	RENI_RAT
44	670.5	32.5	400	1	RENI_SHEEP
45	668	32.4	419	1	KDAP_MOUSE

RESULT 1					
CHYM_BOVIN	STANDARD:	PRT:	381 AA.		
AC P00794:					
DT 21-JUL-1986 (Rel. 01, Created)					
DT 01-JAN-1990 (Rel. 13, Last sequence update)					
DT 16-OCT-2001 (Rel. 40, Last annotation update)					
DE Chymosin precursor (EC 3.4.23.4) (Preprorennin).					
GN Cym OR CPC.					
OS Bos taurus (Bovine).					
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;					
OC Bovidae; Bovinae; Bos.					
OX NCBL_TaxID=9913;					
RN [1]					
RP SEQUENCE FROM N.A. (CHYMOSIN B).					
RX MEDLINE=86301873; PubMed=3091454;					
RA Hidaka M., Sasaki K., Uozumi T., Beppu T.;					
RT "Cloning and structural analysis of the calf prochymosin gene.;"					
RL Gene 43:197-203(1986).					
RN [2]					
RP SEQUENCE FROM N.A. (CHYMOSIN B).					
RX MEDLINE=82221400; PubMed=6283469;					
RA Harris T.J.R., Lowe P.A., Lyons A., Thomas P.G., Eaton M.A.W.,					
RT "Molecular cloning and nucleotide sequence of cDNA coding for calf					
RL prorennin."					
RN [3]					
RP SEQUENCE OF 59-381 (CHYMOSIN B), AND DISULFIDE BONDS.					
RX MEDLINE=79239460; PubMed=381305;					
RA Foltmann B., Pedersen V.B., Kauffman D., Wybrandt G.;					
RT "The primary structure of calf chymosin.;"					
RL J. Biol. Chem. 254:8447-8456(1979).					
RN [4]					
RP SEQUENCE OF 17-77 (CHYMOSIN B).					
RX MEDLINE=7602411; PubMed=240697;					
RA Pedersen V.B., Foltmann B.;					
RT "Amino-acid sequence of the peptide segment liberated during					
RL activation of prochymosin (prorennin).;"					
RN [5]					
RP ACTIVE SITE PEPTIDES OF CHYMOSIN B.					
RX MEDLINE=75060332; PubMed=4612029;					
RA Chang W.-J., Takahashi K.;					
RT "The structure and function of acid proteases. III. Isolation and					
RL characterization of the active-site peptides from bovine rennin.;"					
RN [6]					
RP SEQUENCE FROM N.A. (CHYMOSIN A).					
RX MEDLINE=83054629; PubMed=6183168;					
RA Moir D., Mao J., Schumm J.W., Vovis G.F., Alford B.L.,					
RT "Molecular cloning and characterization of double-stranded cDNA					
RL coding for bovine chymosin.;"					
RN [7]					
RP Gene 19:127-138(1982).					

Query Match 93.4%; Score 1928; DB 1; Length 381;
 Best Local Similarity 98.6%; Pred. No. 5,2e-140;
 Matches 365; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

21 AVTHAAETRIPLKYGSLRKALKERHGLEDFLOKQOYGISSKYSGFGEVASVPLNTNYLD 80
 12 ALSOGAETIRIPLKYGSLRKALKERHGLEDFLOKQOYGISSKYSGFGEVASVPLNTNYLD 71

81 SOYFGKITLGPPOEFYVLFDTGSSDFWVPSITCKSNACKNQRDPKRSSTFQNLGKRL 140
 72 SOYFGKITLGPPOEFYVLFDTGSSDFWVPSITCKSNACKNQRDPKRSSTFQNLGKRL 131

141 SIHGTGSMOGLIGDYDTVYNSIVDIQOTVGLSTQEPDGVFTYAEFDGILGMAVPSLASE 200
 132 SIHGTGSMOGLIGDYDTVYNSIVDIQOTVGLSTQEPDGVFTYAEFDGILGMAVPSLASE 191

201 YSIPFEDMMNRHLVAODLFSVYMDRNGQESMLTGAIDPSYTGSLHWPVTVQOYQWF 260
 192 YSIPFEDMMNRHLVAODLFSVYMDRNGQESMLTGAIDPSYTGSLHWPVTVQOYQWF 251

261 TVDSVTISGVVACGCGQALIDTGTSKLVGPSSDILNIQOAIQATQNOYGFEDIDCNL 320
 252 TVDSVTISGVVACGCGQALIDTGTSKLVGPSSDILNIQOAIQATQNOYGFEDIDCNL 311

321 SYMPVTVFEINCKMPLTPSATVSODGFCSTGFOSENHSHQWILGDVFIREFYVSVPDRA 380
 312 SYMPVTVFEINCKMPLTPSATVSODGFCSTGFOSENHSHQWILGDVFIREFYVSVPDRA 371

381 NNLVGLAKAI 390
 372 NNLVGLAKAI 381

RESULT 2
 CHYM_SHEEP STANDARD; PRT: 381 AA.
 AC P18276;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Chymosin precursor (EC 3.4.23.4) (Preprorennin).
 GN CVM.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 ON NCBI_TaxID=9940;
 RN [1]
 RP MEDLINE=90356410; PubMed=2117748;
 RA Pungekar J., Strukelj B., Gubensek F., Turk V., Kregar I.;
 RT "Complete primary structure of lamb preprochymosin deduced from
 cDNA";
 RL Nucleic Acids Res. 18:4602-4602(1990).
 CC - FUNCTION: CHYMOSIN IS SYNTHESIZED IN THE MUCOSA OF THE STOMACH.
 CC THE ENZYME HYDROLYSES CASEIN TO PARACASEIN.
 CC - CATALYTIC ACTIVITY: Broad specificity similar to that of pepsin A.
 CC Cuts milk by cleavage of a single bond in casein (kappa chain).
 CC - SUBUNIT: MONOMER.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC BUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X53037; CAA37209.1; -;
 CC PIR, S10996; CSMHB.

DR HSSP: P00794; ICMS.
 DR MEROPS: A01.006; -;
 DR InterPro: IPR001969; Asp_protease.
 DR InterPro: IPR001461; Pepsin.
 DR Pfam: PF00026; asp. 1.
 DR PRINTS: PR00792; PEPsin.
 DR PROSITE: PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
 KW zymogen; Signal.
 FT SIGNAL 1 16
 FT PROPEP 17 58
 FT CHAIN 59 381
 FT ACT_SITE 92 92
 FT ACT_SITE 274 274
 FT DISULFID 105 110
 FT DISULFID 265 269
 FT DISULFID 308 341
 FT DISULFID 341 42074 MM; D9903528FA071C47 CRC64;
 SO SEQUENCE 381 AA; 42074 MM; D9903528FA071C47 CRC64;

Query Match 88.6%; Score 1829; DB 1; Length 381;
 Best Local Similarity 93.2%; Pred. No. 1.9e-132;
 Matches 345; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

21 AVTHAAETRIPLKYGSLRKALKERHGLEDFLOKQOYGISSKYSGFGEVASVPLNTNYLD 80
 12 ALSOGAETIRIPLKYGSLRKALKERHGLEDFLOKQOYGISSEYSGFGEVASVPLNTNYLD 71

81 SOYFGKITLGPPOEFYVLFDTGSSDFWVPSITCKSNACKNQRDPKRSSTFQNLGKRL 140
 72 SOYFGKITLGPPOEFYVLFDTGSSDFWVPSITCKSNACKNQRDPKRSSTFQNLGKRL 131

141 SIHGTGSMOGLIGDYDTVYNSIVDIQOTVGLSTQEPDGVFTYAEFDGILGMAVPSLASE 200
 132 SIHGTGSMOGLIGDYDTVYNSIVDIQOTVGLSTQEPDGVFTYAEFDGILGMAVPSLASE 191

201 YSIPFEDMMNRHLVAODLFSVYMDRNGQESMLTGAIDPSYTGSLHWPVTVQOYQWF 260
 192 YSIPFEDMMNRHLVAODLFSVYMDRNGQESMLTGAIDPSYTGSLHWPVTVQOYQWF 251

261 TVDSVTISGVVACGCGQALIDTGTSKLVGPSSDILNIQOAIQATQNOYGFEDIDCNL 320
 252 TVDSVTISGVVACGCGQALIDTGTSKLVGPSSDILNIQOAIQATQNOYGFEDIDCNL 311

321 SYMPVTVFEINCKMPLTPSATVSODGFCSTGFOSENHSHQWILGDVFIREFYVSVPDRA 380
 312 SYMPVTVFEINCKMPLTPSATVSODGFCSTGFOSENHSHQWILGDVFIREFYVSVPDRA 371

381 NNLVGLAKAI 390
 372 NNLVGLAKAI 381

RESULT 3
 CHYM_CALJA STANDARD; PRT: 381 AA.
 ID CHYM_CALJA
 AC Q9N2D2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Chymosin precursor (EC 3.4.23.4) (Preprorennin).
 GN CVM.
 OS Callithrix jacchus (Common marmoset).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
 OC Callitrich.
 ON NCBI_TaxID=9483;
 RN [1]
 RP MEDLINE=20250834; PubMed=10788784;
 RA Kageyama T.;

```
RT "New World monkey pepsinogens A and C, and prochymosins. Purification,
RT characterization of enzymatic properties, cDNA cloning, and molecular
RT evolution."
RU J. Biochem. 127:761-770(2000).
CC -1- FUNCTION: Hydrolyses a variety of proteins.
CC -1- CATALYTIC ACTIVITY: Broad specificity similar to that of pepsin A.
CC Clots milk by cleavage of a single bond in casein (kappa chain).
CC -1- ENZYME REGULATION: Inhibited by pepstatin.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- DEVELOPMENTAL STAGE: Expressed in adult, not neonate-specific as
CC in other organisms.
CC -1- MISCELLANEOUS: The optimal pH is around 2.5.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AB038386; BAA90873.1; -.
CC InterPro; IPR001969; Asp_protease.
CC InterPro; IPR001461; Pepsin.
CC Pfam; PF000026; asp; 1.
CC PRINTS; PR00792; PEPsin.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
CC Hydrolyase; Aspartyl protease; Digestion; Gastric juice; Zymogen;
CC Repeat; Signal.
CC FT SIGNAL 1 16
CC FT PROPEP 17 58 ACTIVATION PEPTIDE (BY SIMILARITY).
CC FT ENZYME 59 381 CHYMOSIN.
CC FT REPEAT 92 102 1.
CC FT REPEAT 274 284 2.
CC FT ACT_SITE 92 92 BY SIMILARITY.
CC FT ACT_SITE 274 274 BY SIMILARITY.
CC FT DISULFID 105 110 BY SIMILARITY.
CC FT DISULFID 265 269 BY SIMILARITY.
CC FT DISULFID 308 341 BY SIMILARITY.
CC SEQUENCE 381 AA; 41896 MW; C5820C74C97BB96B CRC64;
CC -----
Query Match 77.8%; Score 1605; DB 1; Length 381;
Best Local Similarity 81.4%; Pred. No. 2.4e-115;
Matches 301; Conservative 30; Mismatches 39; Indels 0; Gaps 0;
OY 21 AVTHAEITRIPLVYKSLRKALKEHGLEDFLOKQOYGISKXSGFGEVASVPLTNYLD 80
D 12 ALSQASGIVRIPLHKHKSRLRALKEHGLEDFLOKQOYGISKXSGFGEVASVPLTNYLD 71
OY 81 SQYRGKITLGPPEFTVLTFTGSSDFWVPSITCKSNCKNHQRDPKRSSTFOMLGRPL 140
D 72 COYRGKITLGPPEFTVLTFTGSSDFWVPSITCKSNCKNHQRDPKRSSTFOMLGRPL 131
OY 141 SIHGTGSMOGLIGYDFVTAVNIDIOQTGSLSTQEPGDVTVYAFEDLILGMAYSLSASE 200
D 132 SIQGTGSMOGLIGYDFVTAVNIDIOQTGSLSTQEPGDVTVYAFEDLILGMAYSLSASE 191
OY 201 YSIFVFDNMNRHLVAODLFSVYMDRNGQESMLTLGALIDPSYTGSLHWPVTVQOYQWF 260
D 192 YSIFVFDNMNRHLVAODLFSVYMDRNGQESMLTLGALIDPSYTGSLHWPVTVQOYQWF 251
OY 261 TVDSVTISGVVYVACGGQALIDTGSTKLVGPSSDILNIQQAIGATQNOYGEFDIDCNL 320
D 252 TVDSVTISGVVYVACGGQALIDTGSTKLVGPSSDILNIQQAIGATQNOYGEFDIDCNL 311
OY 321 SYMTVYVEIKGKXKPLPSAYTSODOGFCTSGROSENHNSKMLIGDVFIREYTSVDPRA 380
D 312 SSMTVYVEIKGKXKPLPSAYTSODOGFCTSGROSENHNSKMLIGDVFIREYTSVDPRA 371
OY 381 NNVLGLAKAI 390
```

```
Db :|||||||
372 SNLVGLAKAI 381
RESULT 4
PEPA_CALJA STANDARD; PRT; 387 AA.
ID PEPA_CALJA
AC Q9N2D4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Pepsin A precursor (EC 3.4.23.1).
GN PEA.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC Callithrix.
OX NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 16-25, FUNCTION, AND ENZYME
RP REGULATION.
RC TISSUE=Gastric mucosa;
RX MEDLINE=20250834; Pubmed=10788784;
RA Kageyama T.;
RT "New World monkey pepsinogens A and C, and prochymosins. Purification,
RT characterization of enzymatic properties, cDNA cloning, and molecular
RT evolution."
RU J. Biochem. 127:761-770(2000).
CC -1- FUNCTION: Shows particularly broad specificity; although bonds
CC involving phenylalanine and leucine are preferred, many others are
CC also cleaved to some extent.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- MISCELLANEOUS: The optimal pH is around 2.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AB038384; BAA90871.1; -.
CC InterPro; IPR001969; Asp_protease.
CC InterPro; IPR001461; Pepsin.
CC Pfam; PF000026; asp; 1.
CC PRINTS; PR00792; PEPsin.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
CC Hydrolyase; Aspartyl protease; Digestion; Gastric juice; Zymogen;
CC Signal; Phosphorylation.
CC FT SIGNAL 1 15
CC FT PROPEP 16 61 ACTIVATION PEPTIDE (BY SIMILARITY).
CC FT CHAIN 62 387 PEPSIN A.
CC FT MOD_RES 129 129 PHOSPHORYLATION (BY SIMILARITY).
CC FT ACT_SITE 93 93 BY SIMILARITY.
CC FT ACT_SITE 276 276 BY SIMILARITY.
CC FT DISULFID 106 111 BY SIMILARITY.
CC FT DISULFID 267 271 BY SIMILARITY.
CC FT DISULFID 310 343 BY SIMILARITY.
CC SEQUENCE 387 AA; 41563 MW; 7A7968A56846BD CRC64;
CC -----
Query Match 53.7%; Score 1108.5; DB 1; Length 387;
Best Local Similarity 58.6%; Pred. No. 1.9e-77;
Matches 215; Conservative 49; Mismatches 94; Indels 9; Gaps 3;
OY 30 RPLPKKGSIRKALKHEHGLEDFLOKQOYGISKXSGGE-----VASVPLTNYLDSQYEG 85
D 18 KVLIRKKSIRKALKHEHGLEDFLOKQOYGISKXSGGE-----VASVPLTNYLDSQYEG 77
```



```
CC SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC CLEAVAGE
CC VIA AN INTERMEDIATE FORM(S).
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -1- CAUTION: IT IS NOT KNOWN IF THIS IS PEP A-2 OR PEP A-3.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59755; CAA2427.1; -.
DR PIR: S16064; S16064.
DR PIR: S19684; S19684.
DR HSSP: P00790; 1PSN.
DR MEROPS: A01.001; -.
DR InterPro: IPR001969; Asp.protease.
DR InterPro: IPR001791; Laminin_G.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; Pepsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KM Hydroxylase: Aspartyl protease; Digestion: Stomach; Gastric juice;
KW Zymogen; Multigene family; Phosphorylation; Signal.
FT SIGNAL 1 15
FT PROPEP 16 40 ACTIVATION PEPTIDE.
FT PROPEP 41 62 ACTIVATION PEPTIDE.
FT CHAIN 63 388 PEPsin A-2/A-3.
FT ACT_SITE 94 94 BY SIMILARITY.
FT ACT_SITE 277 277 BY SIMILARITY.
FT DISULFID 107 112 BY SIMILARITY.
FT DISULFID 268 272 BY SIMILARITY.
FT DISULFID 311 344 BY SIMILARITY.
FT MOD_RES 130 130 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 388 AA; 41703 MW; 706f7ED50fF59CDD CRC64;

Query Match 52.9%; Score 1092.5; DB 1; Length 388;
Best Local Similarity 56.9%; Pred. No. 3.2e-76;
Matches 210; Conservative 53; Mismatches 97; Indels 9; Gaps 3;

OY 28 ITRIPLYKGSRLKALKHGLLEDFLOKQOYGISKSYSGFEVAVS-----PLTYNLDISOY 83
DB 17 IYKVLVKKSLRRNLSEHGLKDFLKKHNNPASKYFPQAEPTLLIDEOPLENYLDVEY 76
OY 84 FGKTYLCTPPOEFTVLFPTGSSDFMVPSTYCKSNACKWHOFDRPKSSTFQNLGKPLSLH 143
DB 77 FGTIGTIGTPADQFTVIRIDTGSNNLMWPSYCYSLACTNNHNFNQDSSSTYTGSTGYT 136
OY 144 YGTSGMOGILGDTVTYSNVDIOQTGVLSTQEPGDVFTVAEFDGILGMAYPSLASEYSI 203
DB 137 YGTSGMFGILGDTVTYVGAGISDTNQIFGLSTEPGSLFYVAPFGIIGLAVPSISSGAT 196
OY 204 PVFENNAHRLVADDLSTVYMDRNGOE-SMLTLAIDPSTYTGSLHWVRYVOQTMQTV 262
DB 197 PVFENINWOGIVSODLFSVYLSADDSGSVYIFGIDSSYVTGSLMWVPSVEGTMQISV 256
OY 263 DSVYISGVVAVCEGGCAILDTGTSKLVGPSDDLINQOAGANQNGEEDIDCDNLSY 322
DB 257 DSIYINGEALACARCCQAVDTGSLTGLPTSPINATIOSDIGASENSDGEVAVSCSAISS 316
OY 323 MPTVVEINGMKPLTPSATYSODQFCTSGFQS-----ENHSQWILGDIIVIREYVFD 378
DB 317 LPDVFETINGIQYVPVSATYILQSGSCISGFGMDVPTESGELWILGDIIVIREYVFD 376
OY 379 RANNVGLA 387
DB 377 RANNVGLA 385
```

```
RESULT 7
PEPA_MACMU STANDARD; PRT; 388 AA.
ID PEPA_MACMU
AC P11489;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pepsin A precursor (EC 3.4.23.1).
GN PGA.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88313666; PubMed=2900796;
RA Evers M.P.J., Zelle B., Bebelman J.P., Pronk J.C., Mager W.H.,
RT Planta R.J., Eriksson A.W., Frants R.R.;
RL "Cloning and sequencing of rhesus monkey pepsinogen A cDNA.";
RL Gene 65:179-185(1988).
RN [2]
RP REVISION.
RA Zelle B.;
RL Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M20788; AAA36902.1; -.
DR PIR: JT0309; PEMOAR.
DR HSSP: P00790; 1PSN.
DR MEROPS: A01.001; -.
DR InterPro: IPR001969; Asp.protease.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; Pepsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KM Hydroxylase: Aspartyl protease; Digestion: Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation.
FT SIGNAL 1 15
FT PROPEP 16 62 ACTIVATION PEPTIDE.
FT CHAIN 63 388 PEPsin A.
FT MOD_RES 130 130 PHOSPHORYLATION (POTENTIAL).
FT ACT_SITE 94 94
FT ACT_SITE 277 277
FT DISULFID 107 112 BY SIMILARITY.
FT DISULFID 268 272 BY SIMILARITY.
FT DISULFID 311 344 BY SIMILARITY.
SQ SEQUENCE 388 AA; 41696 MW; 97f65E3F6C2A793 CRC64;

Query Match 52.6%; Score 1085.5; DB 1; Length 388;
Best Local Similarity 57.2%; Pred. No. 1.1e-75;
Matches 211; Conservative 51; Mismatches 98; Indels 9; Gaps 3;

OY 28 ITRIPLYKGSRLKALKHGLLEDFLOKQOYGISKSYSGFEVAVS-----PLTYNLDISOY 83
DB 17 IYKVLVKKSLRRNLSEHGLKDFLKKHNNPASKYFPQAEPTLLIDEOPLENYLDVEY 76
```


AC P00790;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pepsin A precursor (EC 3.4.23.1).
GN PGA3 AND PGA4 AND PGAS.
OS Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
ON [1]
RX MEDLINE=6316118; PubMed=6300126;
RA Sogawa K., Fujii-Kuriyama Y., Mizukami Y., Ichihara Y., Takahashi K.;
RT "Primary structure of human pepsinogen gene.";
RL J. Biol. Chem. 258:5306-5311(1983).
RN [2]
RP SEQUENCE FROM N.A. (ISOZYME 5).
RC TISSUE=Placenta;
RX MEDLINE=69233110; PubMed=2714789;
RA Evers M.P.J., Zelle B., Bebelman J.P., van Beusechem V., Kraakman L.,
RA Hoffer M.J.V., Pronk J.C., Mager W.H., Planta R.J., Eriksson A.W.,
RA Frants R.K.;
RT "Nucleotide sequence comparison of five human pepsinogen A (PGA)
RT genes: evolution of the PGA multigene family.";
RL J. Genomics 4:232-239(1989).
RN [3]
RP PARTIAL SEQUENCE OF 1-28.
RX MEDLINE=86059312; PubMed=2415509;
RA Ichihara Y., Sogawa K., Takahashi K.;
RT "Isolation of human, swine, and rat prepepsinogens and calf
RT preprochymosin, and determination of the primary structures of their
RT NH2-terminal signal sequences.";
RL J. Biochem. 98:483-492(1985).
RN [4]
RP SEQUENCE OF 16-100 (ISOZYMES 2, 3; 3A; 4 AND 5).
RX MEDLINE=90130402; PubMed=2515193;
RA Athauda S.B., Tanji M., Kageyama T., Takahashi K.;
RT "A comparative study on the NH2-terminal amino acid sequences and
RT some other properties of six isozymic forms of human pepsinogens and
RT pepsins.";
RL J. Biochem. 106:920-927(1989).
RN [5]
RP SEQUENCE OF 16-68 (ISOZYMES 3 AND 5).
RX MEDLINE=89065108; PubMed=3197840;
RA Foltmann B.;
RT "Activation of human pepsinogens.";
RL FEBS Lett. 241:69-72(1988).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=95392399; PubMed=7663352;
RA Fujiwaga M., Chernata M.M., Tarasova N.I., Mosimann S.C.,
RA James M.N.G.;
RT "Crystal structure of human pepsin and its complex with pepstatin.";
RL Protein Sci. 4:960-972(1995).
CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY, ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; J00287; AAA98529.1; -
DR EMBL; J00279; AAA98529.1; JOINED.
DR EMBL; J00280; AAA98529.1; JOINED.

DR EMBL; J00281; AAA98529.1; JOINED.
DR EMBL; J00282; AAA98529.1; JOINED.
DR EMBL; J00283; AAA98529.1; JOINED.
DR EMBL; J00284; AAA98529.1; JOINED.
DR EMBL; J00285; AAA98529.1; JOINED.
DR EMBL; J00286; AAA98529.1; JOINED.
DR EMBL; M26032; AAA60061.1; -
DR EMBL; M26025; AAA60061.1; JOINED.
DR EMBL; M26026; AAA60061.1; JOINED.
DR EMBL; M26027; AAA60061.1; JOINED.
DR EMBL; M26028; AAA60061.1; JOINED.
DR EMBL; M26029; AAA60061.1; JOINED.
DR EMBL; M26030; AAA60061.1; JOINED.
DR EMBL; M26031; AAA60061.1; JOINED.
DR PIR; A00980; PERU.
DR PIR; A30142; A30142.
DR PIR; B30142; B30142.
DR PIR; A22434; A22434.
DR PIR; PX0023; PX0023.
DR PIR; PX0024; PX0024.
DR PIR; PX0025; PX0025.
DR PIR; PX0026; PX0026.
DR PIR; PX0027; PX0027.
DR PIR; S02663; S02663.
DR PIR; S02664; S02664.
DR PDB; IPSN; 20-APR-95.
DR PDB; IPSO; 20-APR-95.
DR MEROPS; A01.001; -.
DR MIM; 169710; -.
DR MIM; 169720; -.
DR MIM; 169730; -.
DR MIM; 169700; -.
DR InterPro; IPR001969; Asp-protease.
DR InterPro; IPR001791; Laminin.G.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp. 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00441; ASP_PROTEASE; 2.
KW Hydrolase, Aspartyl protease, Digestion, Stomach, Gastric juice;
KW Zymogen, Signal; Phosphorylation; 3D-structure; Polymorphism.
FT SIGNAL 1 15
FT PROPEP 16 62
FT CHAIN 63 388
FT MOD_RES 130 130
FT ACT_SITE 94 94
FT ACT_SITE 277 277
FT DISULFID 107 112
FT DISULFID 268 272
FT DISULFID 311 344
FT VARIANT 28 28 /FTID=VAR_006481.
FT VARIANT 58 58 E-> K (IN ISOZYMES 3A, 4 AND 5).
FT VARIANT 92 92 /FTID=VAR_006482.
FT VARIANT 222 222 V-> L (IN ISOZYMES 4 AND 5).
FT VARIANT 265 265 O-> K.
FT VARIANT 265 265 A-> T /FTID=VAR_006484.
FT VARIANT 353 353 /FTID=VAR_006485.
FT VARIANT 353 353 L-> V.
FT VARIANT 376 376 /FTID=VAR_006486.
FT VARIANT 376 376 D-> E.
SQ SEQUENCE 388 AA; 41977 MW; C9CB89BA08F4D78B CRC64;

Query Match 52.4%; Score 1082.5; DB 1; Length 388;
Best Local Similarity 56.7%; Pred. No. 1.9e-75;
Matches 208; Conservative 52; Mismatches 98; Indels 9; Gaps 3;
30 RPLVYKRSLLKALKEHGLEDFLQROOYGISRSYGFGEVAVS-----PLTNTLDSQYRG 85
19 KVLIRKRSRLRLTSLSEGLLKDFLKNHNPARKYFPQWAPLVDVDEOPLNLTDMEXFG 78

Dd	257	SIIINGETICACSCQAVVDGTSLLAGFTSAISKIQSIGSKMLJGEENISSAIDSL	316
Oy	324	PYVVEEINGKMPLTPSATSQDQCFCITSGFOSEN---HSOKWILGVFIREYSVFDR	379
		I::II II III :: I III I : : :	
Dd	317	PDIEFTINNVQYLPSASATYLTKEDDDCIFDFGMNMDITSYGELMTILLGVFRFYFDVR	376
Oy	380	NNNYGLAKA	389
		I	
Dd	377	ANNOVGIAAA	386
RESULT_13			
PEP4_RABIT			
ID_PEP4_RABIT	STANDARD:	PRT:	387 AA.
P28713;			
Dt	01-DEC-1992 (Rel. 24, Created)		
Dt	01-DEC-1992 (Rel. 24, Last sequence update)		
Dt	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	Pepsin II-4 precursor (EC 3.4.23.1) (Pepsin A).		
OS	Oryctolagus cuniculus (Rabbit).		
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.		
OX	[NCBI_TaxId=9986;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91009127; PubMed=2129536;		
RA	Kageyama T., Tanabe K., Koizumi O.;		
RT	"Structure and development of rabbit pepsinogens. Stage-specific		
RT	nucleotide sequences of cdnas, molecular evolution, and		
RL	gene expression during development.";		
J.	Biol. Chem. 265:17031-17038(1990).		
CC	-1 FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS		
CC	INVOLVING PHENALANINE AND LEUCINE ARE PREFERRED. MANY OTHERS ARE		
CC	ALSO CLEAVED TO SOME EXTENT.		
CC	-1 CATALYTIC ACTIVITY: PREFERENTIAL CLEAVE PHE-, LEU-		
CC	-1 DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE		
CC	THE PREDOMINANT ZMOGENS AT LATE POSTNATAL STAGE.		
CC	HORMONES AND RELATED SUBSTANCES.		
CC	-1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE		
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.		
DR	PIR: D38302; D38302.		
DR	HSP; P00790; IPSN.		
DR	MEROPS; A01.001;-		
DR	InterPro: IPRO01969; Asp_protease.		
DR	InterPro: IPRO01461; Pepsin.		
DR	pfam: PF00026; asp. 1.		
DR	PRINTS: PR00792; PEPsin.		
DR	PROSITE; PS00141; ASP_PROTEASE, 2.		
KW	Hydroxylase; Aspartyl protease; Digestion; Stomach; Gastric juice;		
KW	Zymogen; Signal; Phosphorylation; Multigene family.		
FT	SIGNAL	1	15
FT	PROPEP	16	59
FT	CHAIN	60	387
FT	MOD_RES	129	129
FT	ACT_SITE	93	93
FT	ACT_SITE	276	276
FT	DISULFID	106	111
FT	DISULFID	267	271
FT	DISULFID	310	343
FT	SEQUENCE	387 AA; 42052 MW; 21AD0D7782A89585 CRC64;	

```

Query Match      50.2%  Score 1037:  DB 1,  Length 387;
Best Local Similarity  54.7%  Pred. No. 5,5e-72;
Matches 203:  Conservative  57;  Mismatches 103;  Indels  8;  Gaps  3.

Oy  28  ITRIPLYGKSRLKALKEHGLEDEFLQKOQOYISSKY---SGFGVASYPLNTYLDQSYF  84
      : : : : : | | | | : | | : : : | : : | | | : : | |
Db  17  VAKVPLVRRKSLRNKLIEKGLDQDYLTHTPPARKYFPFKEFANYSFSLNLYDAEYF  76
      : : : : : | | | | : | | : : : | : : | | | : : | |
Oy  85  GKLYIGTPROETVLTLDGSSDFWVWPYSYCKSNACKNHOREDFPKRSSTFQNLGKPLSIHY  144
      | : : : : | | | | : | | : : : | : : | | | : : | |

```

```

Db 77 GTISIGTPPODFVYIFDGTSSNLWVPSTYSSCALCAHAKRENPESSSYQGTSETLSITY 136
Oy 145 GTGSMQIGLIDYDITYVSNIVDIQOTVGLSTDEPQDFVYIAEPDGLGNAYSLSASESIP 204
Db 137 GTGSMGTGLIGDYDYKVGKISIEDTNOIFGLSKTEPEPTGLTFAPFEDGLGLAYSISSSDAP 196
Oy 205 VEDNMNMNHIAADLFSFYSDMRNGOE-SMLTLGALDPEYVYSGLSHMVAVTVQOYWQOFVD 263
Db 197 VEDNMNMNGVLVSQDFSFYSYSSDDEKSGSLVMPGGLDSSYTYGSLMWVSYTEGWTQIMD 256
Oy 264 SVTISVYVAVCEGGCOAILDTGTSKTVLSPSSDLNIQQAIGATONQNGEFDIDCDNLSY 323
Db 257 SVSINGETIACADSCQALVDTGTSLTGTPTSATSNISQISYIGASKNLLGSENIYSCAIDSL 316
Oy 324 PTVYFEINGKMPPLTPSAVYTSODQGFCTISGFOSEN----HSQKTLGCVFIREYSVDR 379
Db 317 PDVITTINGIOYPLIPASAVYILKEDDCTSGLEGMNVDTYTGELMTLIGDVFTRYOFTYDR 376
Oy 380 ANNVLGTAKAI 390
Db 377 ANNQGLAAAV 387

RESULT 14
PEP3_RABIT
ID_PEP3_RABIT STANDARD: PRT: 387 AA.
AC P27822:
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pepsin III precursor (RC 3.4.23.1) (Pepsin A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RE SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Kolwal O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -I- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -I- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, M59237; AAA85370.1; -.
DR PIR, E38302; E38302.
DR HSRP, P00791; 4PEP.
DR MEROPS, A01.001; -.
DR InterPro, IPR001969; Asp_protease.
DR InterPro, IPR001461; Pepsin.
DR Pfam, PF00026; asp_1.
DR PRINTS, PS00792; PEPSIN.
DR PROSITE, PS00141; ASP_PROTEASE_2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.

```

FT SIGNAL 1 15
 FT PROPEP 16 59
 FT CHAIN 60 387
 FT MOD.RES 129 129
 FT ACT_SITE 93 93
 FT ACT_SITE 276 276
 FT DISULFID 106 111
 FT DISULFID 267 271
 FT DISULFID 310 343
 SO SEQUENCE 387 AA: 41969 MW: 15459AC81F3699F CRC64:

Query Match 50.1%; Score 1034; DB 1; Length 387;
 Best Local Similarity 53.8%; Pred. No. 9,4e-72;
 Matches 199; Conservative 57; Mismatches 106; Indels 8; Gaps 3;

CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M59235; AAA85369.1;
 CC PIR: C38302; C38302.
 CC HSP: P00790; IPISN.
 CC MEROPS: A01.001; -.
 CC InterPro: IPR001969; Asp-protease.
 CC InterPro: IPR001461; Pepsin.
 CC Pfam: PF00026; asp. 1.
 CC PRINTS: PR00792; PEPsin.
 CC PROSITE: PS00141; ASP_PROTEASE; 2.
 CC Hydrolase: Aspartyl protease; Digestion; Stomach; Gastric juice;
 CC Zymogen; Signal; Phosphorylation; Multigene family.
 CC
 CC SIGNAL 1 15
 CC PROPEP 16 59
 CC CHAIN 60 387
 CC MOD.RES 129 129
 CC ACT_SITE 93 93
 CC ACT_SITE 276 276
 CC DISULFID 106 111
 CC DISULFID 267 271
 CC DISULFID 310 343
 SO SEQUENCE 387 AA: 42100 MW: 66FC331A3DC75891 CRC64:

Query Match 49.9%; Score 1029; DB 1; Length 387;
 Best Local Similarity 54.2%; Pred. No. 2,3e-71;
 Matches 201; Conservative 58; Mismatches 104; Indels 8; Gaps 3;

CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M59235; AAA85369.1;
 CC PIR: C38302; C38302.
 CC HSP: P00790; IPISN.
 CC MEROPS: A01.001; -.
 CC InterPro: IPR001969; Asp-protease.
 CC InterPro: IPR001461; Pepsin.
 CC Pfam: PF00026; asp. 1.
 CC PRINTS: PR00792; PEPsin.
 CC PROSITE: PS00141; ASP_PROTEASE; 2.
 CC Hydrolase: Aspartyl protease; Digestion; Stomach; Gastric juice;
 CC Zymogen; Signal; Phosphorylation; Multigene family.
 CC
 CC SIGNAL 1 15
 CC PROPEP 16 59
 CC CHAIN 60 387
 CC MOD.RES 129 129
 CC ACT_SITE 93 93
 CC ACT_SITE 276 276
 CC DISULFID 106 111
 CC DISULFID 267 271
 CC DISULFID 310 343
 SO SEQUENCE 387 AA: 42100 MW: 66FC331A3DC75891 CRC64:

Search completed: July 31, 2002, 09:25:11
 Job time: 233 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 09:20:58 ; Search time 31.86 Seconds
(Without alignments)
2117.640 Million cell updates/sec

Title: US-09-643-755b-2
Perfect score: 2064
Sequence: 1 MNFLKSPFYAFELCFQYFV.....REYYSVDRANLVGLAKAI 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1883	91.2	366	6	09N1P5	09N1P5 bubalus bub
2	1760.5	85.3	345	6	027951	027951 bos primige
3	1670	80.9	381	6	09GK11	09GK11 camelus dro
4	1601	77.6	380	6	028950	028950 sus scrofa
5	1423	68.9	379	11	09JUX1	09JUX1 ratus norv
6	1248	60.5	242	6	028075	028075 bos taurus
7	1105	53.5	386	6	09GMV6	09GMV6 canis famli
8	1099.5	53.3	385	13	09DEC4	09DEC4 rana catesb
9	1098	53.2	387	6	09GMV9	09GMV9 suncus muri
10	1090	52.8	387	6	09GMV8	09GMV8 sorex ungu
11	1087.5	52.7	385	6	029080	029080 sus scrofa
12	1087.5	52.7	386	6	09GMV7	09GMV7 rhinolophus
13	1084	52.5	390	6	09GK10	09GK10 camelus dro
14	1081	52.4	384	13	09DEC2	09DEC2 xenopus lae
15	1016.5	49.2	382	13	09PRG9	09PRG9 gallus gall
16	995	48.2	388	6	046524	046524 felis silve

17	970	47.0	388	6	046523	046523 equus zebra
18	962.5	46.6	378	13	09PUR9	09PUR9 pseudopleur
19	943	45.7	376	13	09PUR8	09PUR8 pseudopleur
20	922.5	44.7	387	11	09JUX2	09JUX2 ratus norv
21	918.5	44.5	387	11	09D106	09D106 mus musculu
22	918.5	44.5	387	11	09JKE6	09JKE6 mus musculu
23	899.5	43.6	389	6	09GMV5	09GMV5 suncus muri
24	894	43.3	383	13	09DEC3	09DEC3 xenopus lae
25	891.5	43.2	384	13	091322	091322 rana catesb
26	888	43.0	388	6	09GMV2	09GMV2 oryctolagus
27	880	42.6	392	11	09D1R7	09D1R7 mus musculu
28	878.5	42.6	389	6	09GMV4	09GMV4 sorex ungu
29	874.5	42.4	389	6	09GMV3	09GMV3 rhinolophus
30	863	41.8	387	13	09DDV5	09DDV5 salvelinus
31	859	41.6	383	13	09DE45	09DE45 salvelinus
32	850.5	41.2	389	13	09PMK1	09PMK1 gallus gall
33	849.5	41.2	389	13	09W643	09W643 gallus gall
34	846	41.0	370	6	09TWT1	09TWT1 bos taurus
35	846	41.0	375	6	046500	046500 bos taurus
36	845.5	41.0	387	6	046496	046496 bos taurus
37	843.5	40.9	406	13	P70068	P70068 pagothenia
38	827	40.1	396	11	09D1T2	09D1T2 mus musculu
39	822.5	39.8	389	6	09MYK3	09MYK3 sus scrofa
40	822	39.8	376	6	09TXY4	09TXY4 capra hircu
41	821.5	39.8	389	6	09MYK2	09MYK2 sus scrofa
42	821	39.8	397	6	028546	028546 ovis arles
43	812	39.3	386	6	09TWT8	09TWT8 capra hircu
44	801.5	38.8	399	13	093458	093458 podarcis si
45	792	38.4	402	13	057572	057572 chionodraco

ALIGNMENTS

RESULT	ID	PRELIMINARY;	PRT;	366 AA.
09N1P5	09N1P5			
AC	09N1P5			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	PROCHYMOSIN.			
OS	Bubalus bubalis (Domestic water buffalo).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bubalus.			
OX	NCBI_TaxID=89462;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Batish V.K., Mukhopadhyay U.K., Mohanty A.K., Grover S., Kuipers O.P.;			
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO THE PEPTIDASE FAMILY A1; ALSO KNOWN AS THE			
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.			
CC	EMBL: AF177290; AAF27315.1; -.			
DR	HSSP: P00794; 4CMS.			
DR	MEROPS: A01.006; -.			
DR	InterPro: IPR001969; Asp-Protease.			
DR	InterPro: IPR001461; Pepsin.			
DR	Pfam: PF00026; asp. 1.			
DR	PRINTS: PR00792; PEPsin.			
DR	PROSITE: PS00141; ASP_PROTEASE; 2.			
KW	Aspartyl protease; Hydrolase.			
SO	SEQUENCE 366 AA; 40498 MW; 279BED1258B5F15 CRC64;			

Query Match 91.2%; Score 1883; DB 6; Length 366;
Best Local Similarity 97.8%; Pred. No. 5,7e-150;
Matches 357; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 26 AEIRIRIPLYKKSIRKALKKEHGLEDFLOKQOYGISSSKSGFGEVAVPLTNVYDSQYFG 85
|||||
Db 2 AEIRIRIPCKGKSLRKALKKEHGLEDFLOKQOYGVSSKSGFGEVAVPLTNVYDSQYFG 61

QY	86	KIYLGTPEQEEFTVLEFDGSSDFWVPSITCKSNACKNROEDPRKSSYFONLGRPLSTHYG	145
Db	62	KIYLGTPEQEEFTVLEFDGSSDFWVPSITCKSNACKNROEDPRKSSYFONLGRPLSTHYG	1221
QY	146	TGSMOGLIGDYPTVTSNIVDIQGTGVLSTGEPGDVFTYYAEFDGLGNAVPSLASEYSIPV	2050
Db	122	TGSMOGLIGDYPTAVTSNIVDIQGTGVLSTGEPGDVFTYYAEFDGLGNAVPSLASEYSIPV	181
QY	206	FDNNMNRHLVADLFFSVYMDNNGGSMULTGADIPSVYTGSLHNPVTVQOYQWQFTVDSV	265
Db	182	FDNNMNRHLVADLFFSVYMDNNGGSMULTGADIPSVYTGSLHNPVTVQOYQWQFTVDSI	2411
QY	266	TISGVVVAECGGCAIIDDGTGTSKLVGPSPSDLNLIQQAIGATONQYGEFDIDCDMLSYMPT	325
Db	242	TISGVVVAECGGCAIIDDGTGTSKLVGPSPSDLNLIQQAIGATONQYGEFDIDCDMLSYMPT	301
QY	326	VYFELNGMAYPLUTSATYSQDQGCTSGFQSEHNKSQKMLIGDVFIREYYSVFDPRANLVG	385
Db	302	VYSELNGMAYPLUTSATYSQDQGCTSGFQSEHNKSQKMLIGDVFIREYYSVFDPRANLVG	361
QY	386	LAKAI 390	
Db	362	LAKAI 366	
RESULT	2		
Q27951		PRELIMINARY;	PRT; 345 AA.
AC	Q27951		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE		PROCHYMOSIN.	
OS		Bos primigenius (aurochs).	
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC		Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
OC		Bovidae; Bovinae; Bos.	
OX		NCBI_TaxID=9909;	
RN		[1]	
RP		SEQUENCE FROM N.A.	
RA		Tan S.-Y., Zhang Y.-Y., Liu T.-Y., Liu N.-J., Yang K.-Y.;	
RT		"Nucleotide sequence of bovine prochymosin cDNA and the mechanism of	
RT		cDNA deletion."	
RL		Chin. J. Biotechnol. 5:328-332(1989).	
RN		[2]	
RP		SEQUENCE FROM N.A.	
RA		Xu L.;	
RL		Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.	
DR		EMBL; U19786; AAA73476.1; -.	
DR		HSSP; P00794; 4CMS.	
DR		MEROPS; A01.006; -.	
DR		InterPro; IPR001969; Asp_protease.	
DR		InterPro; IPR001461; pepsin.	
DR		Pfam; PF00026; asp. 2.	
DR		PRINTS; PR00792; PPSIN.	
DR		PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.	
QO		SEQUENCE 345 AA; 38034 MW; 396DECEE2996A1FA CRC64;	

Query Match	85.33;	Score 1760.5;	DB 6;	Length 345;
Best Local Similarity	93.78;	Pred. No. 1e-139;		
Matches 342;	Conservative	1;	Mismatches 1;	Indels 21;
				Gaps
QY	26	AEIRIRPLYKRSKSRKALKEHGLEDFLQKQOYISISSKYSFGFVAVSPLNTLYDSYFG	85	
Db	2	AEIRIRPLYKRSKSRKALK-HGLEDFLQKQOYISISSKYSFGFVAVSPLNTLYDSYFG	60	
QY	86	KIYGTSPQDEETVLEFDGSSDFWFWPSIYCKSNACKNQORDPKRSSFTQNLGRPLSHYG	145	
Db	61	KIYGTSPQDEETVLEFDGSS-----DPRKSSFTQNLGRPLSHYG	100	
QY	146	TGSMOGLIGDYTVAVSNIVDQQTAVGLSTQRPGRVFTYAEDDGLGMAYSLSASESIPV	205	

Query Match	Best Local Similarity	80.9%	Score 1670;	DB: 6;	Length 381;
Matches	309; Conservative	34;	Mismatches	28;	Indels
Db	101	TGSMGILGDTVTAVNSILTDIQQIVFGISTOBPGGVFTYAEEDGILGMAFSLASEYSIPV	165		
Qy	206	FDNMNMRHLVADQDFSVYMDNRNGESMLTLGADIDPSTYTGSLHWPVTVOOYQWQFTVDV	265		
Db	161	FDNMNMRHLVADQDLSTVYMDNRNGESMLTLGADIDPSTYTGSLHWPVTVOOYQWQFTVDV	220		
Qy	266	TISGVVACBEGGCAILDPTGTSKLVGPSISDLINQQAIGATONQYGEFDIDCDNLSTMP	325		
Db	221	TISGVVACBEGGCAILDPTGTSKLVGPSISDLINQQAIGATONQYGEFDIDCDNLSTMP	280		
Qy	326	VFEINGKMYPLTPSAVTSDQDGFCTSGFQSEHNSOKWILGADVIFIREYISVPDRANLV	385		
Db	281	VFEINGKMYPLTPSAVSYSDQDGFCTSGFQSEHNSOKWILGADVIFIREYISVPDRANLV	340		
Qy	386	LAKAI 390			
Db	341	LAKAI 345			
RESULT	3				
Q9GK11					
ID	Q9GK11	PRELIMINARY;	PRT;	381 AA.	
AC	Q9GK11;				
DT	01-MAR-2001 (TREMBlrel, 16, Created)				
DT	01-MAR-2001 (TREMBlrel, 16, Last sequence update)				
DT	01-DEC-2001 (TREMBlrel, 19, Last annotation update)				
DE	CHYMOSIN PRECURSOR (EC 3.4.23.4).				
OS	Camelus dromedarius (Dromedary) (Arabian camel).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.				
OX	NCBI_TaxID=9838;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=GASTRIC MUCOSA;				
RA	Kappeler S.R., Farah Z., Puhan Z.;				
RT	"Camel (Camelus dromedarius) Chymosin and Pepsin as Renneting Enzymes				
RT	for Camel Milk.";				
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.				
CC	- 1 - SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE				
CC	EUKAROTIC ASPARTYL PROTEASES FAMILY.				
DR	EMBL: AJ131677; CAC19554.1; -.				
DR	HSSP; P00794; 4CMS.				
DR	MEROPS; A01.006; -.				
DR	InterPro; IPR001969; Asp_protease.				
DR	InterPro; IPR001461; Pepsin.				
DR	Pfam; PF00026; asp; 1.				
DR	PRINTS; PR00792; PEPsin.				
DR	PROSITE; PS00141; ASP_PROTEASE; 1.				
DR	Aspartyl protease; hydrolase; SIGNAL.				
KM	ASPARTYL protease; hydrolase; SIGNAL.				
FT	SIGNAL	1			
FT	CHAIN	59	381	CHYMOSIN.	
SQ	SEQUENCE	381 AA;	42082 MW;	24BABD57B2E7FDD7 CRC64;	

[illegible]

QY	260	FTVDSVTTISGVVAVNACEGGCCALIDTGTSKLVGPPSSDIINTQALIGATONQXGEFDDICDN	319
Db	251	FTVVDSVTTINGVAVNACGGCCALIDTGTSTVLGPPSSDLIKIOMATIGATENRGGEFVNCGN	310
QY	320	LSYMFVTFVEINGKMYPLTPSPATYSODGFCFTSGFQSEHNSOKMILDDVFIREYVSFDR	379
Db	311	LSMFVTFVFEINGRTRYPLSPSATYSKDDGFCFTSGFGDNNSELMILDDVFIREYVSFDR	370
QY	380	ANNVGLAKAI 390	
Db	371	ANNRGVGLAKAI 381	
RESULT	4		
Q28950			
ID	Q28950	PRELIMINARY;	PRT; 380 AA.
AC	Q28950:		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DE	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	PREPROCHYMOSIN PRECURSOR (FRAGMENT).		
OS	Sus scrofa (pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
OX	NCBI_TaxID=9623;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-GASTRIC;		
RA	Foltmann B., Jensen A.L., Loenblad P., Smidt E., Axelssen N.H.;		
RT	"A developmental analysis of the production of chymosin and pepsin in		
RT	pigs.";		
RL	Comp. Biochem. Physiol. 68B:9-13(1981).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-GASTRIC;		
RX	MEDLINE=96252892; PubMed=8673731;		
RA	Houen G., Madsen M.T., Harlow K.W., Loenblad P., Foltmann B.;		
RT	"The primary structure and enzymic properties of porcine prochymosin		
RT	and chymosin.";		
RL	Int. J. Biochem. Cell Biol. 28:667-675(1996).		
EMBL	U14406; AAB08492.1; -.		
DR	HSSP; P00794; 4CMS.		
DR	MEROPS; A01.006; -.		
DR	InterPro; IPR001669; ASP_protease.		
DR	InterPro; IPR001461; Pepsin.		
DR	Pfam; PF00026; asp. 1.		
DR	PRINTS; PRO0792; PEPSTIN.		
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.		
KW	Signal.		
FT	NON_TER	1	1
FT	SIGNAL	<1	15
FT	PROPEP	16	57
FT	CHAIN	58	380
FT	SEQUENCE	380 AA; 41771 MW; ECAB1B78239D2102 CRC64;	
			POTENTIAL.
			ACTIVATION PEPTIDE (POTENTIAL).
			CHYMOSIN.

[illegible]

QY	200	EXSIEVFNPNMNRHLYADDSYVMDRNGQSM.LTIGAIDPSYTTGSLHWVPVTVQOYQ	255
Db	190	ETVTFVFNPMNRHLYADDLFAVYMSRNDGSM.LTIGAIDPSYTTGSLHWVPVTVMOYQ	249
QY	260	FTVDSVTITSGVVAVCEGCGCAILDTGTSLKVPSSDI.LTIOAIGATONOGCEFDICDN	319
Db	250	FTVDSVTITNGVVAVCNGCGCAILDGTGSLAGPSSDI.LTIOALGABESQGEFDICGS	309
QY	320	LSYMPVTFVEINGKMYPLTPSAVTSQDGFCTSGFQSENHSSQKMLIGDVFIREYVSFDR	379
Db	310	LSMPTVTFVEISGRMYPLPSPATVTDQGFCTSGFQSGSKSQHILGAVFIQETYSVDR	369
QY	380	ANNVLGLAKAI 390	
Db	370	ANNRVLGLAKAI 380	

RESULT	5			
09JUX1		PRELIMINARY;	PRT;	379 AA.
ID	09JUX1			
AC	09JUX1;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	PROCHYMOSIN PRECURSOR (EC 3.4.23.4).			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SD.			
RX	MEDLINE=20139732; PubMed=10673373;			
RA	Kageyama T., Ichinose M., Tsukada-Kato S., Onata M., Narita Y.,			
RA	Moriyama A., Yonezawa S.;			
RT	"Molecular cloning of neonate/infant-specific peptidogens from rat			
RT	stomach mucosa and their expressional change during development.";			
RL	Biochem. Biophys. Res. Commun. 267:806-812(2000).			
DR	EMBL; AJ251688; CAB75983.1; "			
DR	HSSP; P00794; 4CMS.			
DR	MEROPS; A01.006; "			
DR	InterPro; IPR001969; Asp-Protease.			
DR	InterPro; IPR001461; Pepsin.			
DR	Pfam; PF000026; asp; 1.			
DR	PRINTS; PRO0792; PEPsin.			
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.			
KW	Signal: Hydrolase.			
FT	CHAIN	1	16	POTENTIAL.
FT	CHAIN	17	379	PROCHYMOSIN.
SO	SEQUENCE	379 AA;	42434 MW;	49PF50CC759A5ADC CRC64;

[illegible]

Db	251	FTVDRITINDEVAVACGGSPAVLDITGTLALLGPGEDILINIGHAIGACVGGQHDGFDIDCMR	310
QY	320	LSYMPVFEELINGKMYPLTPSAVTSQDGFCTSGFQSEHNSQKMLGDVFTIREYVSFDR	379
Db	311	LNFMPTVFEELINGREFFPLPSASVITNQFQSGSSSGFR--HGSQMWTILGDVFTIREFSVDR	368
QY	380	ANNLVGLAKAI 390	
Db	369	ANNRVGLAKAI 379	
RESULT	6		
ID	028075	PRELIMINARY;	PRT; 242 AA.
AC	028075		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	CHYMOSIN C.		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovidae; Bovinae; Bos.		
OX	NCBI_Taxid=9913;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=FOURTH STOMACH MUCOSA;		
RA	MEDLINE=82189915; PubMed=6804449;		
RX	Nishimori K., Kawaguchi Y., Hidaka M., Uozumi T., Beppu T.;		
RT	"Nucleotide sequence of calf prorennin cDNA cloned in Escherichia		
RT	coli."		
RL	J. Biochem. 91:1085-1088(1982).		
DR	EMBL; J00004; AAA30449.1; -.		
DR	HSSP; P00794; ICMS.		
DR	MEROPS; I01.006; -.		
DR	InterPro; IPR001969; Asp_protease.		
DR	InterPro; IPR001461; Pepsin.		
DR	Pfam; PF00026; asp.1.		
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.		
SO	SEQUENCE 242 AA; 26684 MW; 4EBDB7569FD935B4 CRC64;		
Query Match	60.5%;	Score 1248;	DB 6; Length 242;
Best Local Similarity	97.9%;	Pred. No. 6.7e-97;	
Matches 237; Conservative	2; Mismatches	3; Indels	0; Gaps
QY	149	MOGLIGDTVWYNSNIVDILQQTGVLSTQEPGVFTYAEPFDILGMAYPSLASYSIPVEDN	208
Db	1	MOGLIGDTVWYNSNIVDILQQTGVLSTQEPGVFTYAEPFDILGMAYPSLASYSIPVEDN	60
QY	209	MMNRHLVADOLFVSYMDRNGQESKMLTLGAIDPSYTTGSLHWVPVTVQOYQWFTVDSVTS	268
Db	61	MMNRHLVADOLFVSYMDRNGQESKMLTLGAIDPSYTTGSLHWVPVTVQOYQWFTVDSVTS	120
QY	269	GVVVAECGGCAAILDTGSKLVGPSSDILINIOAIGATQNOVGEPDIDCDNLSTYPTVVF	328
Db	121	GVVVAECGGCAAILDTGSKLVGPSSDILINIOAIGATQNOVGEPDIDCDNLSTYPTVVF	180
QY	329	EINGRMVPLTPSAVTSQDGFCTSGFQSEHNSQKMLGDVFTIREYVSFDRANNLVGLAK	388
Db	181	EINGRMVPLTPSAVTSQDGFCTSGFQSEHNSQKMLGDVFTIREYVSFDRANNLVGLAK	240
QY	389	AI 390	
Db	241	TI 242	
RESULT	7		
ID	09GMV6	PRELIMINARY;	PRT; 386 AA.
AC	09GMV6;		
DT	01-MAR-2001	(TREMBLrel. 16, Created)	

DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
 DE PEPSINOGEN A.
 GN PGNA.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RL (1)
 RN SEQUENCE FROM N.A.
 RA Narita Y., Oda S., Takenaka O., Kageyama T.;
 RT "Phylogenetic position of Insectivora inferred from the cDNA sequences
 of pepsinogen A and C."; Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1, ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC EMBL; AB047246; BAB11752.1; -.
 DR HSSP; P00790; IPSN.
 DR MEROPS; A01.001; -.
 DR InterPro; IPR001969; Asp.-protease.
 DR InterPro; IPR001791; Laminin-G.
 DR InterPro; IPR001461; Pepsin.
 DR Pfam; PF00026; asp. 1.
 DR PRINTS; PR00792; PEPsin.
 DR PROSITE; PS00141; Asp.-PROTEASE; 1.
 KM Aspartyl protease; Hydrolyase.
 SQ SEQUENCE 386 AA; 4151 MW; 5F0D1598322390C5 CRC64;

[illegible]

OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasa S.;
RT "Molecular cloning of pepsinogens in Adult Xenopus laevis and Bullfrogs
Rana catesbeiana.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL: AB045376; BAB20092.1; -
DR HSSP: P00794; 4CSN.
DR MEROPS: A01.001; -
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Aspartyl protease; Hydrolase.
SQ SEQUENCE 385 AA; 41702 MW; 5DC8914FC9CEC603 CRC64;

Query Match 53.3%; Score 1099.5; DB 13; Length 385;
Best Local Similarity 53.9%; Pred. No. 3.9e-84;
Matches 213; Conservative 47; Mismatches 112; Indels 9; Gaps 4;

QY 13 LCFQGYFAVVAHAEITRPIPLYKSKLRKALKEHGLEDPLOKQOYGISKY-SGFGEVA 71
DB 5 LRFG---LVLAIECGVAVSLRKESLARKLRNLGLDGLKHHYNPATKTFPSLAQNS 61
QY 72 SVPLTNLDQYFGKITYLGPPOEFTVLEDTGSSDFWVPSIYCKSNACKNHQRPDKRSS 131
DB 62 GEPLQNYMDIEYFGTISIGTPOSTVFIEDGSSNLWVPVSYCCSPACNHNMFPPQSS 121
QY 132 TFOQLGRPLSHYGTSGNOGLGIDVTYVSNIVDIQOTVGLSTQRPDGYTFYAEFDGILG 191
DB 122 TFOQNTNPVSTIOYGTSGSGELGIDTVQVGNITNQIFGLSQSPSGFLYSPDGIIG 181
QY 192 MAYSLASEYSIPVFNMMNRHLVAQDLFSVYMDRNGOE-SMLTGAIDPSYGLHWV 250
DB 182 LAFSLASLSQATPFVFNMMNRGLIPDLFSYLLSQSGSGSVLEFGVDITSTYTNLMNV 241
QY 251 PTVVQOQWQFTVDSVTSISGVVAVCEGSCAIIIDTGTSLKLVGSSDILNIOAIGATONQY 310
DB 242 PLTAETWQIIVDSISIGGVYIACSGSCSAIVDTGTSILAGSTPIANIQYIYGANQDSN 301
QY 311 GEPLIDCDNLSTPVPVEINGKMYPLTPSAVTSODGFCISGFSQSEN-----HSQKALIG 366
DB 302 GOYVINCNNISNMPVYFTINGVQYPLPASAVYRQSDQCTSGFQAMNLPRTSSGDLWILG 361
QY 367 DVFTREYYSVDRANNLVGLA 387
DB 362 DVFTREYIVVDRANNVYAMA 382

RESULT 9
O9GMV9 PRELIMINARY; PRT; 387 AA.
AC O9GMV9;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PEPSINOGEN A.
GN PGNA.
OS Sorex unguiculatus (House shrew) (Musk shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Insectivora; Soricidae; Suncus.
OX NCBI_TaxID=9378;
RN [1]
RP SEQUENCE FROM N.A.
RA Narita Y., Oda S., Takenaka O., Kageyama T.;
RT "Phylogenetic position of insectivora inferred from the cdna sequences
of pepsinogen A and C.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL: AB047243; BAB11749.1; -
DR HSSP: P00790; 1PSN.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Aspartyl protease; Hydrolase.
SQ SEQUENCE 387 AA; 41576 MW; 7F9F818DD541CE CRC64;

Query Match 53.2%; Score 1098; DB 6; Length 387;
Best Local Similarity 57.1%; Pred. No. 5.3e-84;
Matches 209; Conservative 56; Mismatches 93; Indels 8; Gaps 3;

QY 30 RIPLVYKSKLRKALKEHGLEDPLOKQOYGISKY-SGFGEVASVPLTNLDQYFGK 86
DB 19 KVLVYKSKLRNLKENGLEKDFLAKHNVPASKEYFTFATLADQPLVNTMDMEYFCT 78
QY 87 IYLGTPPOEFTVLEDTGSSDFWVPSIYCKSNACKNHQRPDKSSTFQNLGRPLSHYGT 146
DB 79 IGITPPPOEFTVIFDTGSSNLWVPVSYCCSPACNHNMFPPQSSSTFOSTOTLSTAGT 138
QY 147 GSMOGLIGDVTYVSNIVDIQOTVGLSTQRPDGYTFYAEFDGILGMAVSLASEYSIPV 206
DB 139 GSMTGVLGIDTVQVAGIATNQIFGLSQTEPGSFYLSPEFDILGLAYINLSSGATPVF 198
QY 207 DMMNRHLVAQDLFSVYMDRNGOE-SMLTGAIDPSYGLHWVPTVQOQWQFTVDSV 265
DB 199 DMMNRGLVSDLFVYLLSNQSGSVYIFGIDSSYTYGNLWVPLSEGTWQITVDSI 258
QY 266 TISGVVAVCEGSCAIIIDTGTSLKLVGSSDILNIOAIGATONQYGEFDICDNLSTPMT 325
DB 259 TFMGCAIACSGSCAIVDTGTSILGPNNAIANIOKSIASQANQAMVYSCSSIOSLDP 318
QY 326 VYFEINGKMYPLTPSAVTSODGFCISGFSQSEN-----HSQKALIGDVFTREYYSVDRAN 381
DB 319 IVFTINGIYPLPASAVYIILNOODCTSGFQGMIDPPSGELWILGVFTROYFAVFDKGN 378
QY 382 NLVGLA 387
DB 379 NRVGLA 384

RESULT 10
O9GMV8 PRELIMINARY; PRT; 387 AA.
AC O9GMV8;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PEPSINOGEN A.
GN PGNA.
OS Sorex unguiculatus (Long-clawed shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Insectivora; Soricidae; Sorex.
OX NCBI_TaxID=62275;
RN [1]
RP SEQUENCE FROM N.A.
RA Narita Y., Oda S., Takenaka O., Kageyama T.;
RT "Phylogenetic position of insectivora inferred from the cdna sequences
of pepsinogen A and C.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL: AB047244; BAB11750.1; -
DR HSSP: P00790; 1PSN.
DR MEROPS: A01.001; -
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp. 1.

Oy	28	ITRDLPLYGKSLRKALKEHGLLEDFLQKQOIGISSKY--SOFGEVASVPLTNVINDSOYFG	85
Db	17	IKYKPLVKKKSLKRLNMLBQGLLDYDLTKTHSTNPASKSLIKEAASMMVAQPLENTYDMDEYFG	76
Oy	86	KIYLGTPEQFEFTVFDGSSDFWNPSPISYCKSNACKNROFREDPRKSFQNLGKPLSIHYG	145
Db	77	TIGISTPPQEFETVIEDIGSSLMWMPYSYCSSPACSNINRRFPOQSSSTYYQGTNOKLSTAYG	136
Oy	146	TGSMOGLIGDYDTVTYSNIVDIQOFTVGLSTOEBGVFTYAEFGILGNAPSLASESIPV	205
Db	137	TGSMTGILGYDTVQVGGITTDNQJFGLSETEPGEFLTYAPDGLGLATPSIASSGATPV	166
Oy	206	FDNNMNRRLVAQDLFSYVMMDNNGQ--ESMLTIGALDPSYTGSSLMWVYVYQOIQWQFYVDS	264
Db	197	FDNIMNQGVLSDQLFSYLLSNDQGGSVWMMGGIDSSYFTGNLMWVPLSSETWQIVVDS	256
Oy	265	VTISSGVVACGEGCOALIDPSTSKLVPSPSDILNMQOALGATONQYQGEFDDICDNLSTYMP	324
Db	257	ITMNOQVITACGSGSCAIVDITSTILSGPTNVAISAIGYIGASQANANEMVYSCSAINTLP	316
Oy	325	TVVEFINGKMYPLTPSAVTSODQGFCYSGFOS---ENHSQKWLGDVFTIREYYSVFDRA	380
Db	317	NIVEFINGVQYPLRPSAVAVLDSQQGCTSGFQGMIDIPSSGELMTGLDGVFTIRQFTVYDRG	376
Oy	381	NNIYGLA 387	
Db	377	NNQVGLA 383	

RESULT	13			
09GK10				
ID	09GK10	PRELIMINARY;	PRT;	390 AA.
AC	09GK10			
DT	01-MAR-2001 (TrEMBLrel. 16, Created)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)			
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	PEPSIN A PRECURSOR (EC 3.4.23.1).			
OS	Camelus dromedarius (Dromedary) (Arabic camel).			
OC	Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteleostomi; Cetartiodactyla; Tylopoda; Camelidae; Camelus.			
OX	NCBI_TaxID=9638;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=GASTRIC MUCOSA;			
RA	Kappeler S.R., Farah Z., Puhon Z.;			
RT	"Camel (Camelus dromedarius) Chymosin and Pepsin as Renneting Enzymes			
RT	for Camel Milk.;"			
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE			
CC	ENKAROTIC ASPARTYL PROTEASES FAMILY.			
DR	EMBL: AJ31678; CAC19555.1; "-			
DR	HSSP: P00790; IPSN.			
DR	MEROPS: A01.001; "-			
DR	InterPro: IPR001969; Asp-Protease.			
DR	InterPro: IPR001461; Pepsin.			
DR	Pfam: PF00026; asp. 1.			
DR	PRINTS: PR00792; PPSIN.			
DR	PROSITE: PS00141; ASP-PROTEASE; 1.			
DR	Aspartyl protease; Hydrolase; Signal.			
FT	SIGNAL			
FT	CHAIN	1	63	POTENTIAL.
FT	SEQUENCE	390 AA;	42117 MW;	7A679DD6946D7F8D CRC64;

Query Match	52.5%	Score 1084	DB 6	Length 390
Best Local Similarity	55.2%	Pred No. 8e-83		
Matches	207	Conservative	57	Mismatches 97
			Indels 14	Gaps 5
OY	22	VTHAAETRIPIPLKYSKLKALKEHGLEDFDLOKOOYGISIKY-SGFGVASY----	PLT	76
	16	ITH-----KVPILVKKSSLRKNTLEOGKLDLFIKIHNNHNSLKYFPATSSAANFLDEOPIE		70

Oy	77	NYSBOVQGYLYLCTPPOEFLVETDSSDPWWSYICKSNACKNHOHROPKRSKFTQNL	136
Db	71	NYLDEYFEGTISIPPAONFVFIPTDSSNLMWPSTICSSACNHRNRPNEBSSTIQCT	130
Oy	137	GKPLSIHGTSGMOGILGDPVFAVSNIVDIQQTWGLSTOEPCGFVFAEFDGILGNAPYS	196
Db	131	DELSITIGTSGSMGILIGDPVQVGGISDVNQIFGLSETEPGRSLYAPPDGILGLAPYS	190
Oy	197	LASEXSJPEFDMNMNRHLVVAODLFSVYMDRNGOE-SMLTIGALIDPSSYTGSLHWPVTVQ	255
Db	191	ISSSGGTFVPFDMINDEGLISDLEFSVLLSNDESGSVIIGGIDSSYTGSLMMWPVSVE	250
Oy	256	QYQWQTVDSVTISGVVAVACBEGCGCALIDTGTSKLVGPSSDILNTIQALIGATQNOYGEFDI	315
Db	251	GYWOITVWSTIMEEESIASCGCAIVDTGTSLLAGLPDAISNIOYSIGASEDSYGMWV	310
Oy	316	DCDNLSTYPTVFEFLNCKMYLTLTSATYSQDQGTCTSGFOS---ENISQKMLIGDVFIRE	372
Db	311	SCSSISLSPNIVFEINGQVPLSPSAVILLESDDCTSGFGMDLSSESSELMILGDVEIRO	370
Oy	373	YYSVFDNRANNLVGLA	387
Db	371	YTFVFDNRANNQVGLA	385

```

RESULT 14
09DEC2
ID Q9DEC2 PRELIMINARY; PRT; 384 AA.
AC Q9DEC2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PEPsinogen A.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
[1]
RP SEQUENCE FROM N.A.
RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasa S.;
RT "Molecular Cloning of Pepsinogens in Adult Xenopus laevis and Bullfrog
RT Rana catesbeiana."
RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC BUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL; AB045380; BAB20798.1; -.
DR HSP; P00790; 1PSN.
DR MEROPS; A01.001; -.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF000026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Aspartyl protease; Hydrolase.
KW SEQUENCE 384 AA; 4188 MW; 729DD2E7953D9072 CRC64;

```

[illegible]

0.000000

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 18:18:54 ; Search time 9237.8 Seconds
(without alignments)
8963.856 Million cell updates/sec

Title: US-09-643-755B-3

Perfect score: 3957

Sequence: 1 ctgcaggaatcattgcttacc.....acctaccactaagttacc 3957

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_on: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	3957	100.0	3957	6	AX088021	AX088021 Sequence
2	1551	39.2	3502	8	PHYCSP	J01263 Phaseolus v
3	1353.6	34.2	1415	6	AX343913	AX343913 Sequence
4	1173	29.6	1173	6	AX088019	AX088019 Sequence
5	1149.8	29.1	2970	6	AX252300	AX252300 Sequence
6	900.4	22.8	1305	4	BOVCHYMOB	J00003 bovine chym
7	897.4	22.7	1240	6	AR002347	AR002347 Sequence
8	897.2	22.7	1460	6	E00042	E00042 DNA coding
9	896.2	22.6	1291	6	A15836	A15836 chymosin ge
10	895.8	22.6	2733	6	AR073077	AR073077 Sequence
11	895.6	22.6	1175	6	I04058	I04058 Sequence 5
12	895.6	22.6	2726	6	I08097	I08097 Sequence 5
13	894	22.6	2725	4	BOVCHYMOA	J00002 bovine chym
14	894	22.6	1290	6	E00075	E00075 CDNA encod
15	894	22.6	1291	6	E00075	E00075 CDNA encod
16	893	22.6	1289	6	E00144	E00144 CDNA encod
17	892.4	22.6	1311	6	E00108	E00108 DNA coding
18	892.4	22.6	1460	6	E00295	E00295 CDNA encod
19	890.8	22.5	1269	4	BOVCHYMOA	J00004 bovine chym
20	886.8	22.4	1098	6	E00079	E00079 CDNA sequen
21	886.8	22.4	1098	6	E02341	E02341 CDNA sequen
22	875.4	22.1	2282	6	I08098	I08098 Sequence 1
23	870.8	22.0	1101	4	AF177290	AF177290 Bubalus a
24	839.6	21.2	1292	4	OAPPCXY	X53037 Lamb mRNA f
25	776	19.6	1115	12	SYNPROCA	M22593 Synthetic b
26	753.6	19.0	1094	4	BPJ19786	BPJ19786 Bos primige
27	749	18.9	1172	4	CDRI31677	AJ131677 Camelus d
28	734.4	18.6	1130	4	SS014406	U14406 Sus scrofa
29	718	18.1	1262	9	AB038386	AB038386 Callithr1
30	627.2	15.9	969	6	A02000	A02000 Artificial
31	627.2	15.9	969	6	A06446	A06446 Artificial
32	619.6	15.7	1146	12	SYNPROCA	M31016 Synthetic b
33	605	15.3	1259	10	RNO251688	AJ151688 Rattus no
34	534	13.5	1140	6	AX319837	AX319837 Sequence
35	383.4	9.7	1403	4	AB047246	AB047246 Canis fam
36	382.4	9.7	637	6	E05472	E05472 DNA sequenc
37	378.6	9.6	1161	4	AB047245	AB047245 Rhinolph
38	373.8	9.4	1351	9	AB038384	AB038384 Callithr1
39	370.8	9.4	1381	4	AB047244	AB047244 Sorex ung
40	364.6	9.2	1363	4	PIGPEPA	J04601 Pig pepsi
41	363.8	9.2	1362	4	AB047243	AB047243 Suncus mu
42	363.4	9.2	1221	4	CDRI31678	AJ131678 Camelus d
43	363.4	9.2	1276	5	CHKPSN	D00215 Gallus gall
44	361.6	9.1	1266	4	RABPEP1123	M59235 Rabbit peps
45	357.4	9.0	1353	9	MEPEPA23	X59755 M.fusca

ALIGNMENTS

RESULT 1	AX088021	3957 bp	DNA	linear	PAT 17-MAR-2001
LOCUS	AX088021	Sequence 3 from Patent WO0114571.			
DEFINITION	AX088021				
ACCESSION	AX088021.1	GI:13396949			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
synthetic construct.					
artificial sequence.					
1 (bases 1 to 3957)					
REFERENCE					
AUTHORS	van Rooijen, G., Keon, R.G., Boothe, J. and Shen, Y.				
TITLE	Commercial production of chymosin in plants				
JOURNAL	Patent: WO 0114571-A 3 01-MAR-2001;				
Sembiolys Genetics Inc. (CA)					
location/Qualifiers					
source	1..3957				
FEATURES					
source	1..3957				
ORIGIN	1..3957				
1554..2726					
/note="Figure 2"					
/codon_start=1					
/transl_table=11					

```

/protein_id="CAC34661.1"
/db_xref="GI:13396950"
/translation="MNFLEKSEFAFLFCGOYEAVNHAETTRIPLYKGSLSKALK
EHGLEDLEFKOOKQYISKSYGEGEVSAPLTNLDQSGYGLIYGPDEFLPTPT
GSSDFWVPSTCKSNACKNQRDPKRSSTFQNGKRLSTHYGSGMOGLIGYPTVV
SNVLDIQQTVELSTQEPGDVEFTYAEFGILGMAYPSLASEYSIPVDNMARRHLVAD
LFSYVMDRNGQESMLTLGALDSPSYTSLHNPVYQOYWFYDSTYISGVVAENKML
GCQALIDTGFSTKLGPSSDILNIOALGATGNOYGEFIDCDNISTMPTVYELFGKMK
YPLPSAVTISDQDQSGTSGPQSEHNSKWLIGDVFIREYYSVFPBRANIVGLAAAI"

```

Query Match	100.0%;	Score 3957;	DB 6;	Length 3957;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3957; Conservative	0;	Mismatches	0;	Gaps 0;

0Y	1	ctcgaggaattcatcttgtaacctccagatcatcatatagtgaagagtttggtctctctgcgcg	60
0Y	1	ctcgaggaattcatcttgtaacctccagatcatcatatagtgaagagtttggtctctctgcgcg	60
0Y	1	ctcgaggaattcatcttgtaacctccagatcatcatatagtgaagagtttggtctctctgcgcg	60
0Y	61	tgatttttacctcatttaaaggaggtttccaccctaaagaattcgtatcatctccact	120
0Y	61	tgatttttacctcatttaaaggaggtttccaccctaaagaattcgtatcatctccact	120
0Y	121	ttacttgtaacttaatttccataactcttggltgaaatcacaagcttcgcacaga	180
0Y	121	ttacttgtaacttaatttccataactcttggltgaaatcacaagcttcgcacaga	180
0Y	121	ttacttgtaacttaatttccataactcttggltgaaatcacaagcttcgcacaga	180
0Y	181	tatccctacaattattatttgtaaacncttccaacccgctaaattttatgaagtc	240
0Y	181	tatccctacaattattatttgtaaacncttccaacccgctaaattttatgaagtc	240
0Y	241	ccgtatccttaatgtaagtaacatttccatataatgaaatataattacttaattt	300
0Y	241	ccgtatccttaatgtaagtaacatttccatataatgaaatataattacttaattt	300
0Y	241	ccgtatccttaatgtaagtaacatttccatataatgaaatataattacttaattt	300
0Y	301	agcgttggtagaagacataaagaattattcttattcttcttcataataatglttaata	360
0Y	301	agcgttggtagaagacataaagaattattcttattcttcttcataataatglttaata	360
0Y	361	caataaacaagaattcttaccttaagaagaagattcccatattatattttaaataata	420
0Y	361	caataaacaagaattcttaccttaagaagaagattcccatattatattttaaataata	420
0Y	421	ttatcaaatatttttcaaccacgtaaatcctcataataataagttyttccaagaataa	480
0Y	421	ttatcaaatatttttcaaccacgtaaatcctcataataataagttyttccaagaataa	480
0Y	481	aatttaacctcaattttttatttggaacttcttaagaacaccccgatgacaaact	540
0Y	481	aatttaacctcaattttttatttggaacttcttaagaacaccccgatgacaaact	540
0Y	541	agccatttttcttgaataaaaaaaccaatataatctgtatattttttatatacaatga	600
0Y	541	agccatttttcttgaataaaaaaaccaatataatctgtatattttttatatacaatga	600
0Y	601	aaattcaccaacaacatcttbtgtgtaatttcgaagcaagtcagtacgaaaattct	660
0Y	601	aaattcaccaacaacatcttbtgtgtaatttcgaagcaagtcagtacgaaaattct	660
0Y	661	ataattccatttggaactacggaaggaatcgaagaatctgcttttaacgcggaacat	720
0Y	661	ataattccatttggaactacggaaggaatcgaagaatctgcttttaacgcggaacat	720
0Y	721	cttctaaagtaattttaaataagtaactatataatcaagatttcatatacaaatctcaa	780
0Y	721	cttctaaagtaattttaaataagtaactatataatcaagatttcatatacaaatctcaa	780
0Y	781	tattacttcaaaaaataatagataataataataatactctttttaaatttaagtt	840
0Y	781	tattacttcaaaaaataatagataataataataataatactctttttaaatttaagtt	840
0Y	841	ttattacttcaaaaaataatagataataataataataatactctttttaaatttaagtt	840

Qy	841	aatcttgatatttgtagacttgatttactatctactatglttaaatgltttatagt	900
Db	841	AATGTCGAATTGTGACATATGATTTATTTACTACTATGTTTAAATGTTTATAGAT	900
Qy	901	agtttaagaataantaagtaatgtaagtagtggttagagtglttaaggtttacccaacataaac	960
Db	901	AGTTTAAATTAATATTAATAATATGTAAGAGTGTAGAGTGTACCTTAACATTAAC	960
Qy	961	tataacattatggttgactaattttcataatttcttaatttcttactactttcttggt	1020
Db	961	TATAACATTTATGTGAGACTAATTTTCATATATTTTATGTGTTTACTTTCTTGGT	1020
Qy	1021	atgtaagtcgtaactagaatacagtggtgtgcataggaacctctgtggtcttttggttc	1080
Db	1021	ATGTAAGTCCGTAACAGTAATTAACATGGGTGGCCATGGCACCTGTGTGCTTTTGTTG	1080
Qy	1081	atgcataggtcttgcgcaagaaagaacagaacaaagaaaaagaacaaagaagaga	1140
Db	1081	ATGCAAGGGGTCTTGCCCAAGAAAAGACAAGAACAAAGAAAAAACAACAACAGAGAGA	1140
Qy	1141	caaaaacgaatcacacaacaaactcaaatatgactctgcatcagaatcgcgcggtcc	1200
Db	1141	CAAAAACGCAATCACACAACCAACTCAAAATTAATGTCATGGTGATCAAGATCGCGCGTCC	1200
Qy	1201	atgtaatgctaaatgcatgcaagaacacacggttctaactatgaccttaaatggtctcac	1260
Db	1201	ATGTAATGCTAAATGCGATGCAAGCAACGCGTCTTAACATGCACTTTAAATGGCTCAC	1260
Qy	1261	ccatctcaacccacacaaacacatgtgcttttcttatacgaacaaacctgtt	1320
Db	1261	CCATCTCAACCCACACAACACATGTGCTTTTCTTATCATCACACACAAACCACTGT	1320
Qy	1321	atatattcatctcttcgcgcacacctaatcttcttcaactcaacacacgltcaacctgcat	1380
Db	1321	ATATATTCAATCTCTTCGCGCACTCAATTTCTTCACTTCAACACAAAGTCAACCTGGATA	1380
Qy	1381	tgcggtcatcccatgcgccaatctccatgcatgcatggtccaaacacctctctcttatata	1440
Db	1381	TGCGGTCAATCCCATGCGCCAAATCTCCATGCAATGTTCAACCACTTCCTCTTATATATA	1440
Qy	1441	taccttaataaccttaataataactcaactcttcttatacctcatccatccagaagtacta	1500
Db	1441	TACCTATAAATACCTTAATAATACATCACTCTTCTTCAATCCATCCATCCAGAGTACTA	1500
Qy	1501	ctactctactactacataaiaacccaaccaactacataltcaatactactctactatgact	1560
Db	1501	CTACTCTACTACTAT	1560
Qy	1561	tcccttaagcttcccttcttactacgcttctcttctgttctggttcaatctcgttgcgtta	1620
Db	1561	TCCCTTAAGCTTTCCTCTTCTTACTACGCTTTCCTTGTTCGGTCAATATCTTGTGCTGTTA	1620
Qy	1621	ctaagcgttgttagatcaccccgattctctctacaaagttaagtcctctcgttaagcgc	1680
Db	1621	CTACCGTGTGAGATCACCCGCTTCTCTCTTCAAAAGTTAAGTCTCTCCGTATAGCGC	1680
Qy	1681	tgaagagacatggaactctagaagaactctctgagaacaacagltatgcatcagcaga	1740
Db	1681	TGAAGAGACATGGAACCTCTAGAGACCTCTTGCAAGAAACAACAGTATGGCATCGCAGCA	1740
Qy	1741	agtaactccggtctcgttgaagttgtcagtggtgcacttaccactccttgatagat	1800
Db	1741	AGTAACTCCGCTTCGCTGAGAGTTGCTAGCGCTGCACCTTACCACCTCTGTATGTAAT	1800
Qy	1801	accttgggaagatctactctggnaaccccgctcaaggttcaacggtctctcttatactg	1860
Db	1801	ACCTTGGGAAGATCTACTCTGGAAACCCCGCTCAAGAGTTCAACGGTCTCTTTGATACTG	1860
Qy	1861	gttcccttgactcttggtgttccctctatctactcagaagacaatgaccttgaagaacac	1920
Db	1861	GTTCCCTTGACTTCTGGGTTCCCTCTATCTACTCAAGAGCAATGCGCTGCAAGAACCAAC	1920
Qy	1921	aaagattcgcatacgagaagaatcgltcaaccttcagaacttaggcaaacacctgtctatac	1980

[illegible]

Qy 667 cccattgaacactaggaagtaactgaagatctgctttacatgagacacatctcta 726
|||||
Db 661 CCCATTGACACTCAGGAGTAAGTGAAGATCTGCTTTACATGGAGACACATCTTCTA 720
Qy 727 aagtaatttaataagttactatatacaagatttcatalatacaataactcaatatac 786
|||||
Db 721 AAGTAATTTTAAATAGTACTATATTCAGAAATTTCAATATATCAAAATCAATATTAAC 780
Qy 787 ttctaaaaaatlaattgataataataataactcttttlaatttlaagtttaattgtc 846
|||||
Db 781 TTCGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGT 840
Qy 847 tgaattgtgactatgattattattctactatgatttaatttttataagattta 906
|||||
Db 841 TGAATTTGTGACTATGATTTATTTCTACTATGTTAAATTTGTTTATATGATGTTTA 900
Qy 907 aagtaataataagtaagtagtagagtagtagagtagtagagtagtagagtagtagag 966
|||||
Db 901 AAGTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAC 960
Qy 967 attatgtgtgactaattttcatalatttctatgttcttactctttcttgatgttaa 1026
|||||
Db 961 ATTATGTTGACTTAATTTTCAATATTTCTTATTTGCTTTTACCTTTTCTTGATGTAA 1020
Qy 1027 gtccgtactagaatttacagtgggtgcacgtgacactgtggtcttttgatagat 1086
|||||
Db 1021 GTCCGTACTAGTAATTTACAGTGGGTTGCCATGCACTGTGGTCTTTTGGTCTAGTAT 1080
Qy 1087 ggggtctgcgaagaaaaagacaaagacaaagaaaaagacaaagacaaagacaaagac 1146
|||||
Db 1081 GGGGCTTGGCGAAGAAAAAGCAAAAGAAAGAAAAAGCAAAAGCAAGAGAGCAAAAGAC 1140
Qy 1147 gcaatccacaacccaactcaaatatagtcactgctgctgacaaagtcgcgcgtccatgtat 1206
|||||
Db 1141 GCAATCCACAAACCAACTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200
Qy 1207 gtctaaatgcatgcaagaacacacgctgcttaacatgcaactttaaagtcacacacatc 1266
|||||
Db 1201 GTCTAAATGCCATTCACAAAGCAACACGTCGTTAACATGCACTTTAAATGGCTCCCATCT 1260
Qy 1267 caaccacacacacacacacatgcttcttctctcaacacacacacacacacacacacac 1326
|||||
Db 1261 CAACCCACACACAAACACATTTGCTTTTCTTCAATCAACACACACACACACACACACAT 1320
Qy 1327 tcaattccttcgcgcacactcaatttctcaacttcaacacacgtaacactgcatgtgtg 1386
|||||
Db 1321 TCATTTCTTCCGCCACCTCAATTTCTTCACTCAACACACAGTCAACCTGCAATTAATGCGTG 1380
Qy 1387 tcatcccatgcccacacatctcatalgctgttccacacacacacacacacacacacacacac 1446
|||||
Db 1381 TCATCCCATGCCCAAAATCTCATGATGTTCCAAACACCTTCTCTTAATTAATTAATTAAT 1440
Qy 1447 taaataccttaataataactacttcttcaatccatccatccatccatccatccatccatcc 1506
|||||
Db 1441 TAAATACCTTAATATATCATCTTCTTCAATCAATCAATCAATCAATCAATCAATCAATCA 1500
Qy 1507 tactactataataacccaacccaacacacacacacacacacacacacacacacacacacac 1557
|||||
Db 1501 TACTACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 1551

RESULT 3
AX343913
LOCUS AX343913 1415 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 5 from Patent WO0200899.
ACCESSION AX343913
VERSION AX343913.1 GI:18491959
KEYWORDS
SOURCE Phaseolus vulgaris.
ORGANISM Phaseolus vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
REFERENCE
1 (sites)
AUTHORS Angenon,G., de Jaeger,G., Goossens,A. and Depicker,A.
TITLE Heterologous gene expression in plants
JOURNAL Patent: WO 0200899-A 5 03-JAN-2002;
Vlaams Interuniversitair Instituut voor Biotechnologie vzw. (BE)
FEATURES
source location/Qualifiers
1..1415
/organism="Phaseolus vulgaris"
/db_xref="taxon:3885"
BASE COUNT 469 a 259 c 172 g 515 t
ORIGIN
Query Match 34.2%; Score 1353.6; DB 6; Length 1415;
Best Local Similarity 99.6%; Pred. No. 4.3e-204;
Matches 1367; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
Qy 8 aattacgtactccagatcatcatatagtaagaatttgctctctgcggtgtttt 67
|||||
Db 44 AATTGATTTGACTCCAGATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 103
Qy 68 ttaccctatlaaagggttttccaccataaatctgtaacttcaacttcaacttcaacttca 127
|||||
Db 104 TTACCTGATTTAAAGGGTTTCCACCTAAATAATTCGTGATTCATTTCACTTTACTTG 163
Qy 128 ttaacttaattctcaataacttgggttgaattatcaacgtctccgacacagatccct 187
|||||
Db 164 TTACTTTAATTTCTCATTAATCTTTGTTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 223
Qy 188 acaatttatatttggtaaacatttcaaacggataaaatttataagttcccggtcta 247
|||||
Db 224 ACAATTTATTAATTTGTTAAACATTTTCAACCCGATTAATTAATTAATTAATTAATTAATTAAT 283
Qy 248 tcttaaatgtaactcaactttcatalatgaataataatattacttaatttaagcttg 307
|||||
Db 284 TCTTTAATGTAGTCAATATTTTCAATTTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 343
Qy 308 gtagaagaacataaagatttatcttctctctcatalataaagtlttaataataata 367
|||||
Db 344 GTAGAAACCATTAATTAATTTATTTCTTATTTCTTCAATTAATTAATTAATTAATTAATTAAT 403
Qy 368 aacaaattcttacttaagaagatttcccatlttataattttaaataatatatacaaa 427
|||||
Db 404 AACCAATTTCTTACCTTAAGAAGATTTCCCATTTTAATTAATTAATTAATTAATTAATTAAT 463
Qy 428 atatcttcaacacgtaaatctcatalaataaagttgttccaagaagtaataaatttaa 487
|||||
Db 464 ATATTTTCAACACAGTAAATCTCATTAATTAATTAATTTGTTTCAAAAGTAATTAATTAAT 523
Qy 488 ctccataatttttatttactgactgtacttaagaacacacacacacacacacacacacacac 547
|||||
Db 524 CTCCTAATTTTATTTATTCGACTGATCTTAAAGCAACCCGATGACACAACTACCATCAAT 583
Qy 548 ttttcttltgaataaaaaatccaaattatcattgtattttttttaaagaagaatttc 607
|||||
Db 564 TTTTCTTTGGATTAATAAAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 643
Qy 608 accaaacaatcatgtgtgatttcttgaagaacacacacacacacacacacacacacacacacac 667
|||||
Db 644 ACCAAACATCATTTGTGTGATTTCTGAAAGCAAGCATGTTATGCAAAATTTCTAATTAATTC 703
Qy 668 ccaattgacactaggaagtaacgtaagaactgcttltacatgtagagacacacttctaa 727
|||||
Db 704 CCATTTGACACTAGGAGTAAGTGAAGATGCTGTTTAAATGACGACACACATCTTTAA 763
Qy 728 agtaatttataatagttactatataatgaagatttcaatataataataataataataata 787
|||||
Db 764 AGTAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 823
Qy 788 tctaaaaaatatagatacaataaataataataataataataataataataataataataataata 847
|||||

BASE COUNT	299 a	308 c	262 g	304 t
Db	824	TCTAAAAAATTAATTAGTATTAATTAATAATTAATCTTTTAAATTTTAAAGTTAAATTCCTT	883	
Qy	848	gaatttgcactatgtatattatctactactatglttaaatgtttatagatagtttaa	907	
Db	884	GAATTTGACTAATTTGATTTTATTATCTACTATGTTTAAATGTTTATAGATAGTTTAA	943	
Qy	908	agtaataataagtaatgtaagtagagtggttaagtggttaacccaataactataaca	967	
Db	944	AGTAATATTAAGTAATGATAGATGTTAAGTGTTAACCCTTAAACCAATTAACATAACA	1003	
Qy	968	tttaatgtagaactaatttccatataatcttcatgttccattaccccttctgtatgaag	1027	
Db	1004	TTTATAGTGAGACTAATTTTCAATATTTCTTATTTCTTTTACCTTTTCTGGTATGTAAG	1063	
Qy	1028	tcgcgaactagaattacagtggtgtgcatagtagcctgtggtcctttgtgtcatgcatg	1087	
Db	1064	TCCGTAACATGAAATTAATCTGTGGTGGCATGGCACTCTGTGGTCTTTTGTTGATGATG	1123	
Qy	1088	ggt -cttcgcgaagaaagcaaaagcaaaagcaaaagcaaaagcagagagaaaac	1146	
Db	1124	GATGGTTCCGCAAGAAAAAGACAAAGAACAAAGAAAAAGACAAAGAGAGACAAAAAC	1183	
Qy	1147	gcaatcacacaacccaactcaattagtaactagtcagtcgtatcaagatgcgcgttcaatg	1206	
Db	1184	GCAATCACACAAACCACTCAATTAATGTAAGTACGTGCGTGAATCAAGATGCGCGCTCATGAT	1243	
Qy	1207	gtctaataatgcataagcaaacacagtgcttaacatgtacattaaatggtcacaaccatc	1266	
Db	1244	GTTAAATAGCCATGCMAAAGCAACAGCTGCTTAACATGACCTTAATATGATCACCACATCT	1303	
Qy	1267	caaccacaacacaacacacattgaccttttcttcatcatcaacacaacacactgtatatat	1326	
Db	1304	CAACCACACACAAACAAATGCTTTTCTTTCATCATCATCACCAACCAACCTGATATATAT	1363	
Qy	1327	tcaatcctcttcgcgcaccccaatttcttcaacttcaacaacagtcacactgtca	1378	
Db	1364	TCAATCTCTTCGCGCACCTCAATTTCTTCTTCACTTCAACACAGCTCAACTGCA	1415	
RESULT	4			
LOCUS	AX088019	1173 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO0114571.			
ACCESSION	AX088019			
VERSION	AX088019.1	GI:13396947		
KEYWORDS	.			
SOURCE	Cow.			
ORGANISM	Bos taurus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;			
	Bovidae; Bovinae; Bos.			
	1 (bases 1 to 1173)			
	van Rooijen G., Keon R.G., Boothe J. and Shen Y.			
	Commercial production of chymosin in plants			
	Patent: WO 0114571-A 1 01-MAR-2001;			
	Sembiosys Genetics Inc. (CA)			
	location/Qualifiers			
	1..1173			
	/organism="Bos taurus"			
	/db_xref="taxon:9913"			
	1..1173			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="CAC34680.1"			
	/db_xref="GI:13396948"			
	/translation="MNLKSPPEYAFLCFGYFVAVTHAEITRIPLYKGSLSRLKLM			
	EHGLLEDFIAQQYGISISKSGFGEVAVPLNTLDSYFEGKILGTPPOEFTVLFDP			
	GSDFWVPSIYCKSNACKNHORPFRKSTPQNLGKPLSHIGTSGMGLIGDTVTY			
	SNVLDIOOTVGLSTOEPEDVTVAAEPDILGMAYSLASESIPVEDMMNRHLVAD			
	LFASYVMRNGGSMELTGADIDPSYVSGLHVPVTVQOVMQEPDVSATISGVVVA			
	GCQALIDTGTSKLVGPSDDIINTQAIATATONOGEPDIDCNLSYMFYVFEINGR			
	YPLPSPATTSDDGECFTSGFSESNHSMGLIDVFIREYYSVFDRAANLVGLAAAI			
FEATURES				
SOURCE				
CDS				

[illegible]

DB	Accession	Gene	Protein	Score	DB	Length	Gap
Db	961	ACTGACATGCTGACATGCTGCTTGTAGATGATCAACGGCAGAGATGTACCACTGACCCCTCC	1020				
Qy	2574	gctataccagccagagatcaaggctctgcacaaatgagatccagatgagaaacatcc	2633				
Db	1021	GCCTATACACAGCAGATCAAGGTTCTGCACAGATGATTCAGAGTGAAGAACCATTC	1080				
Qy	2634	cagaaatgagatcttgggagatgcttcatctgttgagtactacaagctcttgcagagcc	2693				
Db	1081	CAGAAATGGATCTTGGGAGATGTGTTCATCTCGTAGTACTACAGCGCTTTTGACAGGGCC	1140				
Qy	2694	aacaacccctgttggctagctaaagcaatctga	2726				
Db	1141	AACAACCTGCTTGGCTACTACTAAGCAATCTGA	1173				
RESULT	5						
AX252300		2970 bp	DNA	Linear	PAT	05-OCT--2001	
LOCUS	AX252300						
DEFINITION	Sequence 2 from Patent WO0168887.						
ACCESSION	AX252300						
VERSION	AX252300.1	GI:15985641					
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
FEATURES							
source							
BASE COUNT	1018 a	547 c	527 g	878 t			
ORIGIN							
Query Match	29.1%	Score 1149.8;	DB 6;	Length 2970;			
Best Local Similarity	99.0%;	Pred. No. 5.4e-172;					
Matches 1157;	Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;			
Qy	2735	aagatgaactaaatgcatgtagtgatgaaagccatgagaaagatggaatctgtatc	2794				
Db	1796	AAGTATGAACATAATGCATGTAGTGTGAAGGCTCATGAGACATGGAATATGTATC	1855				
Qy	2795	cgacacgttaacagtatataaactcgaagctcactctcactctctctatgaaataaacaag	2854				
Db	1856	CGACCATGTAAACAGATATATACTAGGCTCCATCTCATCTTCTTATGATAAACAAGG	1915				
Qy	2855	atgtaataatataaactctatctatgcacctatattgttctatgataaaattccctc	2914				
Db	1916	ATGTTATATATATATTAACACTCTATCTATGACCCCTTATTTGTCATGATAAATTTCTCT	1975				
Qy	2915	tatattataaactcgaagctcgaagcgtctatggaatgcttaaatagtaacaanaac	2974				
Db	1976	TATTTATTAATATCATCTGGAATCGTAGCGCTTATGGAATGCTTCAATATGACAAAAAC	2035				
Qy	2975	aaatgtgactataaagactcttcaacaacttcaactttagcatgtgaaacgagacataa	3034				
Db	2036	AAATGTGATCTATAGACTTCTTAACAATCTTAAGTACTTGAAGTGTGAGACGACATTA	2095				
Qy	3035	gtgttaagaagacataacaatatataatggaagaattgtctccattatataatatata	3094				
Db	2096	GTTTAAAGAGACATTAACAAATTAATGGAAGAAAGTTTGTCCATTTATATATATATA	2155				
Qy	3095	ttaccacattatgattatagagatgtaaagagacataacaattataaagagaagag	3154				
Db	2156	TTACCCACTTATGTTTATATATGAGATGTGAAGGACATTAACAAATTAATGAAGAGAG	2215				
Qy	3155	tttgataccattatataatataactaacattatataatataactataactataactatt	3214				

[illegible]

b, of the enzyme and its precursor are known and a third form seems likely (see bovchymoa, bovchymc). this sequence has tentatively been identified as coding for preprochymosin b. sequence comparison indicates that the precursors for chymosins a and b differ by only two amino acids, and for b and c by only four amino acids.

FEATURES

Location/Qualifiers
source 1..1305
/organism="Bos taurus"
/db_xref="taxon:9913"
mRNA <1..1305
/product="chymob mRNA"
CDS 26..1171
/note="preprochymosin b"
/codon_start=1
/protein_id="AA30448.1"
/db_xref="GI:162860"
/translation="MRCLVLLAVFALISGAEITRIPLKSLRKALKEHLEDEL
OKOYVSISSKSGFEVAVSLPNVNLDSOFKIVTGTPOEFTLVEDSSPEWPS
ITKSNACKNHQFDPKRSSTFONLKRPLSHYGTGSMOCLIXDVTYVNLIVDIOF
VGLSTGPDVFTYAEFDITLGMAYPSLSEISIPVDNMNHHLVADLFVYMDRN
GQESMLTLGLINPSYVTGSLHWVPVTVQOFTVDSVITSGVVAEGGCOALIDLG
TSKLVGPSDILNIOALIGATQMGVEFDIDCNLSMPVVEINGKMPPLNPSAVT
SDQGFCTSGFQSEHNSQKWIIGDVIFREYYSVFDRLNNLVGLAKAI"
sig_peptide 29..73
/note="presequence"
mat_peptide 200..1168
/product="chymosin b"
BASE COUNT 305 a 393 c 340 g 267 t
ORIGIN pst-1 site.

Query Match 22.8%; Score 900.4; DB 4; Length 1305;
Best Local Similarity 87.8%; Pred. No. 1.2e-132;
Matches 982; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 1609 tcgttcgtgttactcgaagctgtgagatccaccgcatccctctctacaaagtaagctc 1668
DB 54 TCTTCGCTCTCCCAAGGCGCTGAGTACACAGGATCCTCTGTCAAAAGCAAGTCTC 113
QY 1669 tcctgaagcgtgaagaagatgagactctagaagactcttgcaagaacaagatg 1728
DB 114 TGAGGAAGCGCTGAAGAGATGGCTTCTGGAAGACTTCTCGAAGAAAGCAAGTATG 173
QY 1729 gcatcagaagaagactcggctcgttgaagttgtagcgtgcacttacaactacc 1788
DB 174 GCATGACGACGAAGTACTCCGGCTTCGGGAGGTGGCCAGCTGCCCTGACCACTAAC 233
QY 1789 ttgatgcaataactttgggaagatctaccctcggaaaccccgcttaagaagtcacgcttc 1848
DB 234 TGGATAGTCAGTACTTGGGAGATCTACCTCGGAGCCCGCCGCAAGGATTCACCGTGC 293
QY 1849 tcttgatgactggttcctcgtgactcgtgtccctctatctactgcaagaaatgctc 1908
DB 294 TGTTTGACACTGCTCTGACTTCTGGGTACCTCTATCTACTGCAAGACATGCGCT 353
QY 1909 gcaagaacacacaagaatcgaatccgagaagtcgltccacttccagaactaagcaaac 1968
DB 354 GCAAAAACACAGCGCTTCGACCCGAGAAAGTCTCCACTTCAGAAACCTGGGCAACG 413
QY 1969 ccttgtctataactaagtgtaagtgatgcaagaaatcttaagctatgataagctga 2028
DB 414 CCGCTCTCATTCACATACGAGGACAGGACAGCATGAGGGCTCTAGGCTATGACACCGTGA 473
QY 2029 cgtgtccacaacttgtagactcaacagacagtaggaacttgagcccaagaacagtg 2088
DB 474 CTGTCTCCAACTTTGGGACATCCAGACAGACAGTAGCGCTGAGCACCGAGACCGCGGG 533
QY 2089 atgtctcaactatgcaagaatcgaatgcatccttgtagtggacataccatcgctcgct 2148
DB 534 ACGTCTTCACTATGCGCAATTCGACGGGATCTCGGGATGGCGTACCCCTCGCTGCGCT 593
QY 2149 cagagtaactgaatacctgtgttgacaacatgataagcaagcaactagtagctaaagct 2208
DB 594 CAGAGTACTGATACCGGTGTTGACAAATGATGAACAGGCACTTGCGCCCAAGACC 653
QY 2209 tgtctcgtgttacaatgagaagaatgagcagagagagcatgctcagcgttgagctatg 2268
DB 654 TGTCTCTGGTTTACATGAGACAGGAATGGCCAGAGACATGCTACCTTG66GCAATCA 713
QY 2269 atccatctactacacagatctctctcactgggtttccagtcactgtgcaagtaactgac 2328
DB 714 ACCGCTCTTACTACACAGGGTCCCTGACATGGGGTCCCGGAGACAGTGCACAGTACTGGC 773
QY 2329 aattcactgtgagacgtgtcaacatcagcaggtgtgtgtgttgatgtgaagtgtgagtc 2388
DB 774 AGTTCACGTGAGACATGTCACCATCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 833
QY 2389 aagatccttggataccggtgacgtgtccaaagtgtgtgcagactagcagagatcttcaaca 2448
DB 834 AGGCACTTTTGGACACGGGCACTTCCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 893
QY 2449 ttcaagcaagctatggaagcacaacagaaacagtaagtgatgtgatatagatgtgcaga 2508
DB 894 TCCAGCAGGCGCATTTGGAGCCACACAGAACAGTACGATGATTTGACATGACTGCGACA 953
QY 2509 accttaactatagcctacagttgtcttggatcaacgycgaagtgtaccactgacc 2568
DB 954 ACTGAGCTACATGCCACACTGTGCTTTGAGATCAATGGCAAAATGACCCACTGACCC 1013
QY 2569 ccttcgctatacagcagcagatcaaggtgtcttcacagatggaattccagatgagaac 2628
DB 1014 CCTCGGCTATACCAAGCCCAAGACGAGGCTCTGTACAGTGCTTCCAGAGTAAAAATC 1073
QY 2629 attccagaatgagatcttggagagatgttcatctcgttgaactacagcgtctttgaca 2688
DB 1074 ATTCCCAAAATGATCTCTGGGGATGTTTTCATCCGAGATTAACAGCTTTTGACA 1133
QY 2689 gggccaacaacctcgttggcgtgaagcaatcga 2726
DB 1134 GGCCCAACACCTGCTGTGGGCTTGCCCAAGCCATCTGA 1171

RESULT 7

AR002347 AR002347 1240 bp DNA Linear PAT 04-DEC-1998
LOCUS Sequence 2 from patent US 5741665.
DEFINITION AR002347
ACCESSION AR002347
VERSION AR002347.1 GI:3963901
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1240)
AUTHORS Kato, E.K. and Stuart, W.Dorsey.
TITLE Light-regulated promoters for production of heterologous proteins
in filamentous fungi
JOURNAL Patent: US 5741665-A 2 21-Apr-1998;
FEATURES
Location/Qualifiers
source 1..1240
BASE COUNT 274 a 374 c 339 g 253 t
ORIGIN

Query Match 22.7%; Score 897.4; DB 6; Length 1240;
Best Local Similarity 87.4%; Pred. No. 3.7e-132;
Matches 982; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 1609 tcgttcgtgttactcgaagctgtgagatccaccgcatccctctctacaaagtaagctc 1668
DB 100 TCTTCGCTCTCCCAAGGCGCTGAGTACACAGGATCCTCTGTCAAAAGCAAGTCTC 159
QY 1669 tcctgaagcgttgaagaagatgagactctagaagactcttgcaagaacaagatg 1728
DB 160 TGAGGAAGCGCTGAAGAGCATGGGCTTCTGAGAGTCTCTGCAAGAAACAGCATATG 219

Qy 1729 gcatcagcagcaagctacccgctcgttgaaagtctagcgtgcacttaacactacc 1788
|||||
Db 220 GCATAGCAGCAAGTACTCCGGCTTCGGGGAGTGGCCAGCCTGCCCTCGACCACTTAC 279
|||||
Qy 1789 ttgatagcactactcttgggaagatctcctcgcgaaccccgctcaagaagtctc 1848
|||||
Db 280 TGGATAGTCACTACTTGGGAAGATCTACCTCGGAGACCCGCCGACGAGTTACCGTGC 339
|||||
Qy 1849 tctttgatacgtgtccctctactcttgggtccctctatctactgcgaagaaatgctt 1908
|||||
Db 340 TGTTTGACACTGGCTCCTCTGTGACTTCTGGGTAACCTCTATCTACTGACAGCAATGCT 399
|||||
Qy 1909 gcaagaaccacaaagaattcgaatccgagaagctgcacactccacgaactaggcaac 1968
|||||
Db 400 GCAAAAACACACAGCGCTTCGACCCGAGAAATGCTCCACTTCCAGAACTCGGGCAAC 459
|||||
Qy 1969 cctgtctactactcgaatggaatgacatgacgaagaactcttagctatgatacgtca 2028
|||||
Db 460 CCTGTCTATCCACTACACGAGGACAGCAGCAGCATGCAAGGCACTCTAGGCTATGACAC 519
|||||
Qy 2029 ctgtctccaacattgtgacatcttcaacagacatgagacttagaccccaagaacagtg 2088
|||||
Db 520 CTGTCTCCAAATGTGGACATCCAGAGACATAGGCTTGACACCAAGAGCCGGGG 579
|||||
Qy 2089 atgtctcactatgcaagaattcgaatgacatcctgtgtatgacatccatcgcgcgt 2148
|||||
Db 580 ACGTCTTACCTATGCGCAATTCGACGGGATCTGTGGGATGGCCTTACCCCTCGCTGCT 639
|||||
Qy 2149 cagaatctcgaatcctgtgttgaaacatgatagaacacgaactagtagctcaagact 2208
|||||
Db 640 CAGATATCTCATACCCCTGTTTGACAAATGATGAACAGGACCTGCTGGCCCAAGAC 699
|||||
Qy 2209 tttctcgtgttatacagcagaatgacaggaagacatgtcactcagcttgagatctg 2268
|||||
Db 700 TGTCTCTCGTTTAAATGACACAGAAATGGCCAGAGAGCATGCTACGCTGGGGGCAATCG 759
|||||
Qy 2269 atccatctactacacagaatctctcacttggttccagtcactgtgcagcagtagctgc 2328
|||||
Db 760 ACCGCTCTACTACACAGAGGTCCTGCGACCTGGTCCGCTGACAGTGCAGCATCTGCG 819
|||||
Qy 2329 aattcactgtgagacagtgtaacacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgt 2388
|||||
Db 820 AGTTCTACTGTGACAGTGTACATCAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 879
|||||
Qy 2389 aagctatcttgatcacccgttaccagcgtgtgcagcagcagcagcagcagcagcagc 2448
|||||
Db 880 AGGCAATCTCTGACACGGGCACTCTCAAGCTGTGGGCCAGGACGACATCTCAACA 939
|||||
Qy 2449 ttcagcaagctatgagacacacagaacagcagcagcagcagcagcagcagcagcagc 2508
|||||
Db 940 TCCAGCAGGCGATTGGAGCCACACAGAAACGATGAGATGATGATGATGATGATGATG 999
|||||
Qy 2509 accttaagctacatgctcagcagctgtcttgatgaacacagcagcagcagcagcagcagc 2568
|||||
Db 1000 ACCGACTCTATATCCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1059
|||||
Qy 2569 ccttcgcctatcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2628
|||||
Db 1060 CCTCGCCTATACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1119
|||||
Qy 2629 attcccaagaatgagcttgggaagatgtgttcattcgtgtaactacacagcgtcttgaca 2688
|||||
Db 1120 ATTCCAGAAATGATCTCTGGGGGATGTTTCATCCGAGATATATACCGCTCTTTGACA 1179
|||||
Qy 2689 gggcgaacaactcgttggcgttagcctaagaacatcgaagctt 2731
|||||
Db 1180 GGGCAACAACTCTGTGGGCTGTGGCCAAAGCAGCATCTGACTGCT 1222
|||||

RESULT 8
E00042 E00042 1460 bp RNA Linear PAT 29-SEP-1997
LOCUS
DEFINITION DNA coding of pre-pro renin.

ACCESSION E00042
VERSION E00042.1 GI:2168348
KEYWORDS JP 1982141287-A/1.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1460)
AUTHORS Banardet, R.A., Jien, M., Donarudo, T.M., Arison, T.R. and Jierardo, F.B.
TITLE RENNIN, PREPRORENNIN OR PRORENNIN GENE OBTAINED FROM RECOMBINED DNA MATERIAL AND LIVE CELL CONTAINING GENE
JOURNAL Patent: JP 1982141287-A 1 01-SEP-1982;
COMMENT KOBORATEIBU RES INC
OS calf
PN JP 1982141287-A/1
PD 01-SEP-1982
PE 14-JAN-1982 JP 1982003556
PR 16-JAN-1981 US 81 225717, 01-DEC-1981 US 81 325481 PI
BAANDET, RABUTSUTSUKII ARUFUODO, JIEN MAO, PI DONARUDO
TEIRAA MOJIA
PI ARISON TAUNTON RIGUBIT, JIERARDO FURANSHISU BUOBUSU PC
C12N1/00, C07G/00, C07H21/04, C12N15/00, C12P21/00//C12N1/18, PC
C12P19/34, C12R1/19, C12R1/865;
PC strandedness: Double;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue_type=stomach; Location/Qualifiers
FH Key
FH CDS
FT 205..1350
FT /product=pre-pro renin/
FT mat_peptide
FT 205..1347
FT /product=pre-pro renin/
FT CDS
FT 253..1350
FT /product=pro renin/
FT mat_peptide
FT 253..1347
FT /product=pro renin/
FT CDS
FT 278..1350
FT /product=pro renin/
FT mat_peptide
FT 278..1347
FT /product=renin/
FT /product=renin/
FEATURES
source location/Qualifiers
1..1460
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 328 a 440 c 398 g 294 t
ORIGIN

Query Match 22.7%; Score 897.2; DB 6; Length 1460;
Best Local Similarity 87.7%; Pred. No. 3.9e-132;
Matches 980; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

Qy 1609 tctgt 1668
|||||
Db 233 TCTTGTCTCTCTCCAGGGCGCTGAGATCACAGAGATCTCTGTGAAGGCAAGTCTC 292
|||||
Qy 1669 tccgtgaagcgtggaagaacatgactctcagaagaactcttgagaagaacagatg 1728
|||||
Db 293 TGAGGAAGCGCTGAGAGGATGGGCTTCTGAGAGATCTCTGAGAAACAGCAGTATG 352
|||||
Qy 1729 gcatcagcagcaagctacccgctcgttggaagtgtgtagcgtgcacttaacactacc 1788
|||||
Db 353 GCATAGCAGCAAGTACTCCGGCTTCGGGAGGTGGCCAGCGTGCCTTGACCAACTACC 412
|||||
Qy 1789 ttgatagcactacttgggaagatctaccccggaacccgctcaagaagttcacgttc 1848
|||||
Db 413 TGGATAGTCACTTGTGGGAAGATCTACCTCGGAGACCCGCCGACGAGTTACCGTGC 472
|||||
Qy 1849 tcttgaacatggttccctctactcttggttccctctatctactcgaagcaagcagct 1908
|||||

Db	473	TGTTTACACATGGCTCTCTGACTTCTGGGTACCCCTATCTACTGACAGCAATGCT	532
Qy	1909	gcaagacacacacaaagattcgatccgagaagtcgctccacctccagaaattagcgcaac	1968
Db	533	GCAGAAAACCAACGCGCTTGCACCCGAGAAAGTCGTCCACTTCCAGAACCTGGCGAAGC	592
Qy	1969	ccttgcctatacactacgtagtagtagtagtagtagtagtagtagtagtagtagtagtag	2028
Db	593	CCCTGCTATACCACTACGGGACAGGACGATGACGAGGACCTTGGGCTATGACACCTCA	652
Qy	2029	ctgtcccaacactgtgtgagatctcaacacagtagtagtagtagtagtagtagtagtag	2088
Db	653	CTGTCTCCACATTTGTGACATCCAGACAGATGAGGCTTGAGACACCCAGAGCCCGGG	712
Qy	2089	atgtcttcacctaagcagaattcgatggatcctgtgtatggataccatcgctcgt	2148
Db	713	ACGCTTTCACCTATGCGGAATTTGAGCGGATCTGCGGATGGGCTACCCCTCGCTCCCT	772
Qy	2149	cagaagtagctgatacctgtgtgtgaacaatgaaacgacacactagtagtagtagtag	2208
Db	773	CAGAGTACTGATACCGCTGTTTGACAAATGATGAACAGACACCTGCTGGCCCAAGACC	832
Qy	2209	tgttctcgtgttcaatgagaagaatgacagagagacatgctcagctgtgagtagtag	2268
Db	833	TGTTCTCGGTTTACATGAGACAGAAATGCGCAGAGACATGCTACCTGGGGCCATCG	892
Qy	2269	atccatcciaactacacagagatctctcactgtgtgtccagtagctgtgtgacagtagctg	2328
Db	893	ACCGCTCTACTACACAGAGGCTCCCTGCACGTGGGTGCCGTGACAGTGCAGCAGTACGTGC	952
Qy	2329	aattcaactgtgagacagtagtcaacatcagcaggtgtgtgtgtgtgtgtgtgtgtgtgt	2388
Db	953	AGTTCACCTGTGACAGTGTCAACATCAGCGGTGTGTTGTGCTGTGAGGCTGGCTGTC	1012
Qy	2389	aagatctctgtgagacagtagcgtccaaagctgtgtcgaacctgacagcagacattccaaca	2448
Db	1013	AGGCAATCTCGGACACGGGCACTCTCAAGCTGTGCGGGCCACGACGACACTTCTCAACA	1072
Qy	2449	ttcagcaagctatttgaagccacacagaaacagtagcgtgtgtgtgtgtgtgtgtgtgtgt	2508
Db	1073	TCCAGCAGGCGCATTTGGAGCCACAGAAACAGTACGATTTGATGATTCAGATCGGACA	1132
Qy	2509	accttaagctatactgactagtagtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2568
Db	1133	ACCTGAGCTATCATGCCCCACTGTGTTTGTGATCAATGGGAAATGTACCCACTGACCC	1192
Qy	2569	ccttcgacatcacagcagcagagatcaaggtgtgtcgaacctgtgtgtgtgtgtgtgtgtgt	2628
Db	1193	CCTCGGCTATACAGGACGACAGGAGGCTTGTGACAGTGGCTTCCAGATGAAATTC	1252
Qy	2629	attccagaagaatgagatcttggagagatgtgtcattcgttgaatctacacagcgtcttgaaca	2688
Db	1253	ATTCCAGAAATGTGATCTGTGGGGATGTTTTCATCCGAGATATATACAGCGTCTTTGACA	1312
Qy	2689	gggccaacaacactcgt	2726
Db	1313	GGGCCAACACCTCGTGGGCTGGCCAAAGCCATCTGA	1350
RESULT	9		
LOCUS	A15836	1291 bp	DNA
DEFINITION	chymosin gene.		linear
ACCESSION	A15836		
VERSION	A15836.1		
KEYWORDS	GI:488959		
SOURCE			
ORGANISM	Bos taurus		
REFERENCE	1 (bases 1 to 1291)		
AUTHORS	Simons, A.F.M. and De Vos, W.M.		

TITLE DNA fragments, containing a lactic acid bacterium-specific regulator region for the expression of genes coding for normally heterologous proteins

JOURNAL Patent: EP 0307011-A 5 15-MAR-1989;

FEATURES NEDERLANDS INSTITUUT VOOR ZUIVELONDERZOEK

source 1. .1291

location/Qualifiers

BASE COUNT 323 a 382 c 328 g 257 t 1 others

ORIGIN

Query Match 22.6%; Score 896.2; DB 6; Length 1291;

Best Local Similarity 87.6%; Pred. NO. 5.7e-132;

Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

Qy	1609	tcgttgcgttactacacgcctgctgagatcaaccgcctctctctacaaagtagtctc	1668
Db	9	TCTTCGCTCTCTCCAGGGCGCTGATCAACAGATCCCTGTACAAAGCAAGTCTC	68
Qy	1669	tcgtaagcgctgaagaacatgtagctctagaagactcttcgagaacttcgagaacagtagt	1728
Db	69	TGAGGAAGCGCTGAAGAGATGGCTTCTGAGGACTTCTCAGAAACAGCAGTATG	128
Qy	1729	gcatacagagaagtagctccggtctgggtgaagtgtgtgtgtgtgtgtgtgtgtgtgtgt	1788
Db	129	GCATCACCAGCAAGTACTCCGGCTTGGGAGGTGGCCAGGTCCTCTGACCAACTAC	188
Qy	1789	ttgataagtaactcttggagaagatctaacctcggaaacccgcctcaagaagtagctctc	1848
Db	189	TGATATGCACTACTTTTGGAGATCTACCTCGGGACCCGCCAGAGATTCACCGTGC	248
Qy	1849	tcttgaabactgttctcctcgtactctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1908
Db	249	TGTTGACACTGGCTCCCTGACTTCTGGGTACCTCTATCTACTGCAAGCAATGCT	308
Qy	1909	gcaagaacacacaaagattgataccgagaagtagtgcacacttcagaacttagcaaac	1968
Db	309	GCAGAAACACACGCTTCCAGCCGAGAAAGTGTCCACTTCCAGAACTGGGCAAGC	368
Qy	1969	ccttgcatacactacagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag	2028
Db	369	CCCTGTCTATCTACCTACGAGACAGGACATGACGAGGATCTGTGGCTATGACACCGTCA	428
Qy	2029	ctgtctccaaactgtgtgacatcaacagcagtagtagtagtagtagtagtagtagtagtag	2088
Db	429	CTGTCTCCAACTGTGTGACATCCAGCAGACAGTATGAGCTTATGACACCCAGAGCCGGG	488
Qy	2089	atgtcttaccatagcaaatctgtagtagtagtagtagtagtagtagtagtagtagtagtag	2148
Db	489	ACGTTTACCTATGCGCAATTCGACGGGATCTGGGATGGCTTACCCCTCGCTCGCT	548
Qy	2149	cagaatctgatactctgttctgaacaatgataagacacacacacacacacacacacacac	2208
Db	549	CAGATCTACTGATACCGCTGTTTGTGACAAATGATGAACAGGACTGTGTGGCCCAAGCC	608
Qy	2209	tgcttcggttatactgacaggaatggtgcagagagacatgctcaagctgtgtgtgtgtgt	2268
Db	609	TGTTCTCGGTTTACATGAGACAGGAATGCGCAGGAAGCATGCTCAAGCTGGGGCCATCG	668
Qy	2269	atccatctctacacacagagatctctcactggttcagtagtagtagtagtagtagtagtag	2328
Db	669	ACCCNTCTACTACACAGGCTCCCTGCACTGGGTCCCGGTGACAGTGCAGCATGCTGCG	728
Qy	2329	aattcaactgtgacagtagtcaacacacagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2388
Db	729	AGTTACTGTGACAGTGTGACATCAGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	788
Qy	2389	aagcatctttagataccgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag	2448
Db	789	AGGCAATCTGTGACACAGGCACTTCAAGCTGTGTGGGCCAGCAGCATCTCAACA	848

Best Local Similarity 87.6%; Pred. No. 7.2e-132;
Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY	1609	tcgttgtgttactccaagctcgctcgagataaccgcattccctctcaaaagttaagctc	1668
Db	49	TCCTTCGGCTCTCTCCACAGGGGCTGAATACCCAGATGCTCTGTATCAAAAGCAAGTCTC	108
QY	1669	tccttaagcgctbaaagaacatggagctcttaaaagactctctgcagaacaacagatg	1728
Db	109	TGAGGAAGGCCCTTAAGGACATGGGCTTGTGAGGACTTCTCGAGAAACAGCATGTG	168
QY	1729	gcaticgaagcaagtlactccggtctcgltgaagltgctagcgltgcacttaaccaatc	1788
Db	169	GCATCAGACGCAATACTCCGGCTTGGGGAGGTGSCAGCTGCCCTTGACCAACTAC	228
QY	1789	ctgtagtgaataactttgggaagaatctactcggaaaccccgctcaagaagttaacgttc	1848
Db	229	TGGATATGCACTACTTTTGGGAAGATCTACCTCGGGACCCCGCCAGAGATTCACTG	288
QY	1849	tcctttgaatcagtgctctctcgactctcggttccctctatctacgtcaagaagatgct	1908
Db	289	TGTTTGACACTGGCTCCTCTGACTTCTGGGTACCCCTCTTACTGCAAGACATGGCT	348
QY	1909	gcaagaacaccacaaagatctgatalccgagaagaatcgltccaccccttcagaacttaagcaac	1968
Db	349	GCAAAAACACACACGCTGTGACCCGAGAAAGCTGCCACTTCAGAACTGGGCAAGC	408
QY	1969	ccctgtctatacactaagtagcagatagatctgaagaagatcttaagctatgatacgtca	2028
Db	409	CCCTGTCTATCCACTTACGGGACAGGACATGTAGAGGCACTCTTAAGTATGACACCTGA	468
QY	2029	ctgtctccaacatctgttgacatccaagaacagtaagacttagaacccaagaacaggtg	2088
Db	469	CTGTCTCCAAATTTGTGGACATCCAGAGAGACATTAAGCCTTGAGCACCAAGAGCCGGG	528
QY	2089	atgtctcaaccatlgcagaatctgatalgcctctgtglatggcataccatcgctcgct	2148
Db	529	ACGCTTTCACCTTAAGCCGATTTGACGGGATCTCGGGATGAGCTTACCCCTGCTCGCT	588
QY	2149	caagagtctcgtataccgtgtttgaaacaatatagaaccgcaacttagtctaagaagt	2208
Db	589	CAGGTACTCGATNCCGCTTTTGAACAATGATGAAACGGGCACTGGTGGCCCAAGCC	648
QY	2209	tgcttcggtttacatbgaagaatctggccagaagagcaatgtcaagctttgaagctatg	2268
Db	649	TGTTCTCGGTTTACATGGACAGGAATGGCCAGSAGAGCATGCTCAGCCTGGGGGCGCATGC	708
QY	2269	atccatccctactacaagaatctctctcaacttggttccagtaactgtgcagaagtaagtc	2328
Db	709	ACCCGCTCTACTACACAGGGTCCCTCAGCTGGGTGCCCTGACATGACAGCACTAGTCGC	768
QY	2329	aattcacttggaagagtgatcacataagcggtgtgtgttgcatagtgtaagtgatgc	2388
Db	769	AGTTTACTGTGAGAGGTATACCATCAAGCGAGTGTGTGTCCTGTATAGGTGGCTGTCTC	828
QY	2389	aagtalcttggataccggtacgttccaagctgttcgaaactgcagctgcagacatctccaca	2448
Db	829	AGGCAATCTGGACACGGGCACTCCAACTGAGCTGGGGCCAGCAGCAATCCTCAACA	888
QY	2449	ttcagcaagctattggagccacaagaacacgtacgtgtgattgacatagattggcga	2508
Db	889	TCCAGCAGGGCATATGGAGCCACACAACACAGTACATGATTTGACATGCACTGCGACA	948
QY	2509	accttaactcaatgctcaagctgtctttggagatacaagcgcaagatgtgtaaccatgtacc	2568
Db	949	ACCTGACTCTCATATGCCCATGCTGGTCTTGTGATCATATGCAAAATGTATCCACTGACCC	1008
QY	2569	ccttcgcctatcacagccagatccaaggtttctgcacccagtgatcttcagaagtgaacc	2628
Db	1009	CTCTCGGCTTACCAAGCACAAGACCAAGGCTTCTGTACCAAGTGGCTTCCAGATGAATTC	1068
QY	2629	attcccaaaatgatacttgggaagatgttcatctgtagtactacaagcttcttgaca	2688

Db	1069	ATTCCAGAAATGAGTCTCGGGATGTTTTCATCCGAGAGTATTACACGCTTTTGACA	1128
Qy	2689	gggccacaacactcgttltggtcgtacgtcaaacatctga	2726
Db	1129	GGGCCAACACCTCTGTGGGCTGCCCAAGACCATCTGGA	1166

RESULT	12		
108097			
LOCUS	108097	2726 bp	linear
DEFINITION	Sequence 5 from Patent EP 0301669.		PAT 02-DEC-1994
ACCESSION	108097		
VERSION	108097.1	GI:589192	
KEYWORDS			
SOURCE	Unknown.		

REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
1 (phases 1 to 2726)	van den Berg, J. A. D. and Brake, A. J. D.	DNA constructs containing a Kuyvetromyos alpha-factor leader sequence for directing secretion of heterologous derivatives	Patent: EP 0301669-A1 5 01-FEB-1989;	location/manifestations

FEATURES	Location/Qualifiers
source	1. .2726
BASE COUNT	/organism="unknown"
ORIGIN	751 a 654 c 572 g 749 t

Query Match 22.68; Score 895.6; DB 6; Length 2726;

Best Local Similarity 82.7%; Pred. No. 6.6e-132;

```
Matches 1024; Conservative 0; Mismatches 214; Indels 0; Gaps 0;
```

QY	1610	cgctgcgtgtaactcaagctctcctctgaagatacccgcaatctctctctcaagaagttaagttctc	1669
Db	669	CGATGCTTCCCATCATATATGTGGCTAGATATCAACAGGATCCCTGTGTCAAAAGGCAAGTCTCT	728
QY	1670	ccgttaagaagcctgaaggaatactgaagactcttgaagaactctctgcagaacaagaagatag	1729
Db	729	GAGGAAGGCGGTGAAGAGGATGATGGCTTCTGGAGGACTTCTGCGAAGAAACAGCAATATGG	788
QY	1730	catcagacgaagaagtaaccgcgccttcggttgaagttgcctgaagcaacttaccacttaact	1789
Db	789	CATCAGCAGCAAGTACTCCGGCTTTCGGGGAGGTGGCAGACGCTCCCTGTACCACACTACT	848
QY	1790	tgaatagtaactacttcttggagaagatactactcctcgaaacccgcctcaagaagttaccgyltct	1849
Db	849	GGACAGTCAGTACTTGTGGAAAGATATACCTCGGGACCCCGCCAGAGTTACACGGTCT	908
QY	1850	ctttgataactcgtgttctg	1909
Db	909	GTTTGGACACTGGGCTCCCTGCTACTTCTGGGGTACCCCTTACTACTTCAAGACCAATGTCTG	968
QY	1910	caagaacaccacaagaagtctgaatcccgagaagaagtcctccactctcgaagaactttagcaaac	1969
Db	969	CAAAAAACCAACAGCGCTTGACCCCGAAGAAAGTGTCTCCACTTCCAGAACTGGGCGAAGCC	1028
QY	1970	ctctgtcataactactcgttaacagtagtaactgaagaagactcttaggatactgaataccgtaac	2029
Db	1029	CCTGTCTATCTCCTACTACCGGAGCAGGACGCAATGCAAGGGCACTCTGGGCTATTGACACCGTAC	1088
QY	2030	tgtctccaacatctgttgacaacttaacaagaacagtagaactttagcaaccaagaagaacagtga	2089
Db	1089	TGTCTCCAACTTGTGSACTATCCAGAGACAGTATAGGCTCTGACACCAACCGAGAGCCGGGGA	1148
QY	2090	tgtcttaccatactgaagaattctgaatggcatccttgytlatgycataaccatcgtctcgtc	2149
Db	1149	CGTCTTCACTATGCGCAATTTGACGAGGATMCTGGGGATGGGCTTACCCCTGCTGCGCTC	1208
QY	2150	agagtaactcgtataactcgtgttttgaacaacatgtatgaacgcaactctagtactcgaagact	2209
Db	1209	AGACTACTGCATACCGGTGTTTGCAACACTGTATGAACAGGCACTCGTGGGCCCAAGACT	1268

FEATURES	Location/Qualifiers
OY 2210	gtctccggtttacatgagacagaagaaatggccagaagagatgcctcaagcttggagctattgga 2265
Db 1269	gtttcttcgggtttacatgagacagaagaaatggccagaagacagcttcacgcttggggccatgcga 1328
OY 2270	tccatccctactacacagagatctctcactcaggggtttccagctcactgtgcgaacagttactgga 2329
Db 1329	cccgttccttactacacagagatccctcactcaggggtttccagctcactgtgcgaacagttactgga 1388
OY 2330	attcactgtgagcaatgtcaccatcagcgggtgttggttgttcgatcattgaaagtgtatgta 2389
Db 1389	gtttcactgtgagcagatgctcaccatcagcgggtgttggttgttggtcgtgtgagaggtggcgttca 1448
OY 2390	agctactcttgatatacccgtaagtcgaagcttgatcggaacctagcagcgacatctcacaat 2449
Db 1449	ggcccttctctgtgacacagggggacacctcccaagctgtgtcgggccacacagcagatcctcaacat 1508
OY 2450	tcaagaacatctattgagcgacacacaacagacaggttgatgttgacatagatttcggagaa 2509
Db 1509	ccagcagagccatttgagagccacacagacagatgagctgattgtgacatgcactgcgacaa 1568
OY 2510	cccttagctacatgctcctacagttgtctctttgagaaccaagcagaatgtaaccaactgacccc 2569
Db 1569	cctgagctacatgagccacacagctgctgttggatgtacatgscanaaattgtaccactgaccccc 1628
OY 2570	ctcgcgctatataccagccagatcaaaagggtcttcgacacagttgattccagagtgtgaacca 2629
Db 1629	ctcccccttatccacacacagcagacacagggcgttcgtaccagctggtccacagagcaaaatca 1688
OY 2630	tccccaagaatgagatctctggagatggtgtctatctgtgatctcctcagcgtctttggacag 2689
Db 1689	ttcccaagaatgagatctctggagatggtgtctatctgtgatctcctcagcgtctttggacag 1748
OY 2690	ggccacaacacctcgtctggcgtcgtcgaagaacaaatcgaagcttaataaagtatgaactaaaa 2749
Db 1749	ggcccaacaacacctcgtggggcgtggccaaagacatctgactcagcttggttgaacaacggttg 1808
OY 2750	tgcagtgtggtgtgaagagctcatgtagagagcatgtaatattgataccagcatgtatacagt 2809
Db 1809	ccaaagcctttaaagtgatatttactttaaagctcttcgaattttaaataaattttcttttatatgc 1868
OY 2810	ataataactgagctcactatcctcctctctatagaataa 2847
Db 1869	ttttatgacttgaattgttcaatttatatatttttataatga 1906
RESULT 13	
BOVCHYMOA	
LOCUS	BOVCHYMOA 1275 bp mRNA linear MAM 26-APR-1993
DEFINITION	bovine chymosin a (rennin) mRNA.
ACCESSION	J00002
VERSION	J00002.1 GI:162857
KEYWORDS	chymosin; chymosin A; rennin.
SOURCE	bovine (calf) cdna of fourth stomach mucosa mRNA.
ORGANISM	Bos taurus
REFERENCE	Maniatis, Metaxas; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; 1 (bases 1 to 1275).
AUTHORS	Moir, D., Mao, J. I., Schumm, J. W., Vovis, G. F., Alford, B. L. and Taunton-Rigby, A.
TITLE	molecular cloning and characterization of double-stranded cDNA coding for bovine chymosin
JOURNAL MEDLINE COMMENT	Gene 19, 127-138 (1982) 83054629
FEATURES	chymosin (rennin) is the major proteolytic enzyme in the fourth stomach of the unweaned calf. two chromatographically distinct forms are known and a third seems likely (see bovchymob and bovchymoc). this sequence has been tentatively identified as chymosin a; it differs from chymosin b by only two amino acids and from chymosin c by another amino acid. [1] argues that the different chymosins are probably polymorphic variants of a single chymosin gene.

	source	1..1275	/organism="Bos taurus"	
	CDS	21..1166	/db_xref="taxon:9913"	
			/note="preprochymosin a"	
			/codon_start=1	
			/protein_id="AA30447.1"	
			/db_xref="GI:162858"	
			/translation="MRCVLVLAFAVPALSSQGAETIRIPLYKSGLSRLAKLEHGLLEDPI OKROGSISSKYSGFEVASVPLNTYIDISGYKITLGPDPFYLPTGGSDFWPVR LYCKSNACKNHDRPDPRKSSTFGONRKLSTHYGMSGQILIGDTPVNSNLDIQK VELSTOEPEDVFITAEFDGILGMATPSLASSTLFEDNNMRHLVADLTSVMDTR GOESNLTLGAIDPCYTGSLEHWPVTVQWQVDFSYTISGVVAECBGQDALDTG TSKDGSPSDDLINIOAIGATONOMDEDDICDNLSYMPAVFLFNMGMYDLFTSAYSY SDQDFCTSGFSESHSQMKILIGDVIFIREYVSVPFRANNLVGLAKAI"	
	sig_peptide	24..68	/note="presencequence"	
	mat_peptide	195..1163	/product="chymosin a"	
	BASE COUNT	293 a 391 c 336 g 255 t		
	ORIGIN	20 bases upstream from codon 1.		
	Query Match	Best Local Similarity 22.6%; Score 894; DB 4; Length 1275;		
	Matches 978; Conservative	0; Mismatches 140; Indels 0; Gaps 0;		
Qy	1609	tgttgctgtaactgaagcgttgtgatatacccgatctctcttacaagaagtgaatgc	1668	
Db	49	tcttcgtctctccccaggcgctgaaatpaccaagatccctctgtaacaaagcaatcttc	108	
Qy	1669	tccgtlaaggcgtlgaagagacatgtagcttctaagaagactcttcgacaacaacagtatg	1728	
Db	109	tggaggaaggcgctgaaagagacatgggcttctggagaccttccgcagaacaacagcatag	168	
Qy	1729	gcatacgacgaagtaactccggcgttcgttgaagttgttagcgtgccacttaccnaattac	1788	
Db	169	gcatacgacgaagtaactccggcgttcgttgaagttgttagcgtgccacttaccnaattac	228	
Qy	1789	ttagatgtaaatgaactttggaagaatcttactctggaaaccccgcgtcaagaagttcacg	1848	
Db	229	tggatrtgtaagtaattgggaagatrttaccttccggagacccccgcgtcaagaagttcacg	288	
Qy	1849	tcttgatacttgcttccttgacttcgttgcttggttcccttacttactgcaagaagatgctt	1908	
Db	289	tgttttacacctggctcctcttgacttttggtgtracctcttatctgcaagacatgctt	348	
Qy	1909	gcaagaaccaccaagaatcgatccgagaagaatcgatccacttccaagaacttaaggcaaac	1968	
Db	349	gcaaaaaacaccacagcgcttcgcacccgagaaaatgcgncacacttccagaaacttggcgaac	408	
Qy	1969	ccttgctctataactcaggtacaggtatagatgaagaagaatctttagctatgtataccgtta	2028	
Db	409	ccctgctctatccactacacggagacagcgacatccaggcgcttctggcttatgacacccgta	468	
Qy	2029	cctgtcccaacattgtgascattccaacagacagagtaagtaagtaagcaccacaacagagtg	2088	
Db	469	ctgtctccacacttctcgacatccagcacagacagtaagccttgagcacccacagagccgggg	528	
Qy	2089	atgcttcaactatgcaagaattcgatgtagatcccttgltatgtgacatacccatgctgcgt	2148	
Db	529	acgtcttccactatggcgaattcgacggagatcctggggatggcgtaacccctgcgcgcct	588	
Qy	2149	cagagtaactcagatacctgtgcttgaacaaatgatgaacccaacactagtagctcaagact	2208	
Db	589	cagagtaactcagatacctgtgcttgaacaaatgatgaacccaacactagtagctcaagact	648	
Qy	2209	tgtcttcagtttacaagacagaatgagccagagagacatgtctcaacgcttggagcatatg	2268	
Db	649	tgtcttcagtttacaagacagaatgagccagagagacatgtctcaacgcttggagcatatg	708	
Qy	2269	atccatccactacagaagatccttcaacttgggttccagctacgttgcagagtaactggc	2328	

[illegible]

FEATURES	FT	CDS	26. .1081
source	Location/Qualifiers		
	1. .1290		
	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
BASE COUNT	309 a 384 c 334 g 263 t		
ORIGIN			
Query Match	22.6%; Score 894; DB 6; Length 1290;		
Best Local Similarity	87.5%; Pred. No. 1.3e-131;		
Matches 978; Conservative	0; Mismatches 140; Indels 0; Gaps 0;		
QY	1609	tcgtgtcgtgttactcaacgtctgtgagatcaccccgatcttcctctacaaagttaagtc	1668
DB	54	ttctgtgcctctctcccaaaagccctgagatccacagatccctctgtgacaaaggcaagtctc	113
QY	1669	tcctgaaggcgtctgaaggagcatgtggccttcttgagagacttcttcgaagaacaagatg	1728
DB	114	tcaggaaggcgtctgaaggagcatgtggccttcttgagagacttcttcgaagaacaagatg	173
QY	1729	gcatacagcagcaagtactccgcgtcttcgtgaaagtgtgtagcgltgcacttacaactacc	1788
DB	174	gcatacagcagcagatnctccggcttccggagagtgggccagcgccgccctgcacaaactacc	233
QY	1789	ttagatagcaataactcttgggaagatcactcctggaaaccccgctcaagaagttcacggttc	1848
DB	234	tgatgtgtagtgcagctttgggaagatcttaccctggggaccccccgcagagttgacacgttc	293
QY	1849	tccttgatactgttccctctgactctctgtgtgtccctctacactgacagcaatgct	1908
DB	294	tgctttgacactggcctctctgacttcttgaggtaccccttcttactgacagcaatgct	353
QY	1909	gcaagagaccacaagaatctgcataccgagaagaatgcgtccacactccgaacttaggcaaac	1968
DB	354	gcacaaaccacacagcccttgcagaccagaaagtcgacacacttccgaacttgggcgaacg	413
QY	1969	cctgtctctatacactacggtacaggtagcatgaagaatcttagctatgatacgtca	2028
DB	414	ccctgtcactacactctgagacagcagagatgcaggggcatcttgaagatgtgacacagctta	473
QY	2029	ctgtcccaacatctgtggaacatcaacagacagtagacttagcacccaagaacacaggtg	2088
DB	474	ctgtctcccaacatctgtggaacatcaacagacagtagacttagcacccaagaacacaggtg	533
QY	2089	atgtcttcaactatgagaatctgcatactccctgtgtatgataccaaccatgcctcgct	2148
DB	534	acgtcttcaactatgagaatctgcatactccctgtgtatgataccaaccatgcctcgct	593
QY	2149	cagaagtaactcgatacctgtgttgacaacactgtatgaaacccagtaagctccaaagact	2208
DB	594	cagaagtaactcgatacctgtgttgacaacactgtatgaaacccagtaagctccaaagact	653
QY	2209	tggtctcgtgttaccatggaacgaagaatgagccaggaagagcatggtccacagcttgagactatg	2268
DB	654	tggtctcgtgttaccatggaacgaagaatgagccaggaagagcatggtccacagcttgagactatg	713
QY	2269	atcatcctcaactcaacagagatctcttcaactggtgttccagtaactgtgcaagtaactgac	2328
DB	714	accctgtcctcaactcaacagagatctcttcaactggtgttccagtaactgtgcaagtaactgac	773
QY	2329	aattcactgtggaacagtgtaaccaatcaagcgggtgtgtgtgttgcatgtgaagtgatgtc	2388
DB	774	agttcactgtggaacagtgtaaccaatcaagcgggtgtgtgtgtgttgcatgtgaagtgatgtc	833
QY	2389	aagctactcttggaataccggtacgttcaagctgtgtcggaacttagcaggaacacttctaca	2448
DB	834	agggcactcttggaataccggaacgttcaagctgtgtcggaacttagcaggaacacttctaca	893
QY	2449	ttcgaagactatcttggaacacaacagaacagatcagtgtagtltgacatagattgcaca	2508
DB	894	ttcgaagactatcttggaacacaacagaacagatcagtgtagtltgacatagattgcaca	953

QY 2509 accctagctaacatgctcctacgttgcctttgagatcaacggaagatgtaaccactgacc 2568
|||||
Db 954 ACCTAGCTACATGCCACTGTGCTTTGAGATCAATGGCAAAATGTACCCACTGACCC 1013
QY 2569 cctcgctataccggaagatacaaggcttcgtcaccaagtgatctccagaatggaacc 2628
|||||
Db 1014 CCTCGCTATACCAAGCAAGAGCTTCTGTACCAAGTGCTTCCAGAGTGAATGC 1073
QY 2629 attcccaaatgagctcttggaagatggttcattcgtgagactacagcgtccttgaca 2688
|||||
Db 1074 ATTCCCAAAATGATCTCTGGGGGATGTTTCATCCGAGAGATTATACAGCGCTTTGACA 1133
QY 2689 gggccaacaacctcgttggagtaagtaagaatctga 2726
|||||
Db 1134 GGGCCAACAACCTCGTGGGGCTGGCCAAAGCATCTGA 1171

RESULT 15
A15633
LOCUS A15633 1291 bp mRNA linear PAT 18-FEB-1994
DEFINITION preprochymosin.
ACCESSION A15633
VERSION A15633.1 GI:491951
KEYWORDS
ORGANISM
SOURCE synthetic construct.
synthetic construct
artificial sequence.
1 (bases 1 to 1291)
REFERENCE
AUTHORS Carey,N.H., Doel,M.T., Harris,T.J.R., Lowe,P.A. and Emtage,J.S.
TITLE A process for the production of a polypeptide
JOURNAL Patent: EP 0068691-A 29 05-JAN-1983;
CELLTECH LIMITED

FEATURES
source Location/Qualifiers
1..1291
/organism="synthetic construct"
/db_xref="taxon:32630"
26..1171
/gene="preprochymosin"
26..1171
CDS
/gene="preprochymosin"
/codon_start=1
/transl_table=1
/protein_id="CA01244.1"
/db_xref="GI:491952"
/translation="MRLIVYLAVFALISQAEIRIPLYGKSLRKALKHEGLLEDFL
OKOOGISSKYSGRGVAYPLTNVLDISOYRGKTYLTGPEEYVLEDTGSSDWPVS
IYKSNACKNHORFDPKRSSTFONLGPPLSHIGTGMOGILGDTVVSNIVDIQOT
VGLSTOEPGDFVFAEFDGILGMAYPSLASYSIPVDNMNRLVAQDLFSYMDRI
GQESMLTLAGINPSYTGSLHWYVTVQOOWOFTVDSTVTSIGVVAVAGEGCAILDIG
TSKLVGSSDILNIOALIGATONOGEDIDCNLSYMPVTFEINGKMPYLPFSAVT
SODOGFTSGFSGNSHOSKMLIDGVFIREYVSVDNRANNLVGLAKAI"
misc_difference 1291
/note="POLY A oligo C"

BASE COUNT 309 a 384 c 334 g 263 t 1 others
ORIGIN

Query Match 22.6% Score 894; DB 6; Length 1291;
Best Local Similarity 87.5% Pred. No. 1.3e-131;
Matches 978; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 1609 tcgtgtcgttactcaacgctgctagatcacccgcatcctctctacaagtaagtctc 1668
|||||
Db 54 TCTTGCGCTCTCCCAAGGGCTGAGATCACAGGATCCCTCTGTACAAAGGCAAGTCTC 113
QY 1669 tcgttaagcgctgaagaacatgactctagaagactcttcgagaacaacagtatg 1728
|||||
Db 114 TGAGGAAGGCCCTGAAGGAGCATGGGCTTCTGAGAGACTCTCGAGAAAGCAGATATG 173
QY 1729 gcatacagcagaatgactcggcttcggtgaagtgtcagtgctgcaacttcaactacc 1788
|||||
Db 174 GCATACGACGAAGTACTCTCGGCTTCCGGGAGGTGGCCAGCGTCCCTGACCAACTAC 233

QY 1789 ttgatatgaacttctgggaagatctactcgtgaaccccgctcctaagaatgctacgcttc 1848
|||||
Db 234 TAGATAGTACATCTTTGGGAAGATCTACTCGGGAGCCCCGCCCAAGAGTTTCACTGTC 293
QY 1849 tctttgatactggttcctctcgtacttccttggttcctctatactatgcaagaatgctc 1908
|||||
Db 294 TGTTTGACACTGGCTCTCTGTGACTTCTGGGTACCCCTATCTATGCAAGAGCAATGCTCT 353
QY 1909 gaaagaaccccaagaatctcgatccgagaagatgctcaacttccagaacttagcaaac 1968
|||||
Db 354 GCAAAACCCACAGCGCTTGACCCGAGAAAGTGTGCAACTTCCAGAACTTGGGCAAGC 413
QY 1969 ccttgtctatacctacggttacaggttagcatcaagaagaaacttgaactatgataccgta 2028
|||||
Db 414 CCTCTGCAATCATCTACAGGAGACAGGACAGATGCAAGGCAATCTTAGATATGACACCGTTA 473
QY 2029 ctgtctcaacaattgtggaacatcaacagacagtaggacttagcaaccaagaacagtg 2088
|||||
Db 474 CTGTCTCCAAACATTTGTGGACATCCAGACAGTAGGCTTGACACCCCAAGAGCCGGGG 533
QY 2089 atgtctcaactatcagaatcgcattgcatctctgtatggtacataccactgctcgt 2148
|||||
Db 534 AGCTTTCACTATGCGCAATTCGACGGGATCTGAGGATGGCCCTACCCCTGCTGCTCT 593
QY 2149 cagagtaactcgaatacctgtgttgacaacatgatgaaccgaacacttagtaagtaagta 2208
|||||
Db 594 CAGAGTACTCGATACCCCGTGTGTGACACATATGATACAGGCACTTGGTGGCCCAAGAC 653
QY 2209 tgtctcgtttacatggaacgaagatggaagagaagacatgctcaacgcttggagctatg 2268
|||||
Db 654 TGTTCCTGGTTTACATGACAGAGATGTCAGAGAAAGCATGCTCACCGCTGGGGGCGATCA 713
QY 2269 atccactctacacacagatctctcaactggttcacagtaactctgtgaacagtaactg 2328
|||||
Db 714 ACCCGTCTCTACTACACAGGGTCTCTGACATGGGTGCGCTGACATGACGATCTGCTG 773
QY 2329 aattcactgtgacagctgttcacacacacagcagcgtgtgtgttctcactgtgaagtgatg 2388
|||||
Db 774 AGTTCACTGTGACAGTGTGACACATCATCAGCGGTGTGTAGTGCCCTGTAGAGGTGCTGT 833
QY 2389 aagctactcttgataccggtacgctcacaagctgtgtcgaacttagcagcgacatctcaaca 2448
|||||
Db 834 AGGCGATCTTGGACACAGGGCACCTCCAAGCTGTGCGGGCCAGCAGCATCTCAACA 893
QY 2449 ttcaacagctatttgagacccaacagaacagtaagtgatgttgaactagattcgaca 2508
|||||
Db 894 TCCACAGGCGCATTTGGACCCACACAGAACCAAGTACGGTGAAGTTGACATCGACTGCGACA 953
QY 2509 accctagctatacgtcctacagctgtccttgagatacgaagcaagaatgtaaccactgaccc 2568
|||||
Db 954 ACCTAGAGTACTAGTCCCACTGTGTCTTTGAGATCAATGGCAAAATGTATCCACTGACCC 1013
QY 2569 cctcgctataccagcagaagatacaagggttcgtcaccaagtgatctccagaatggaacc 2628
|||||
Db 1014 CCTCGCTATACCAAGCAAGAGCTTCTGTACAGTGTGCTTCCAGAGTGAATGC 1073
QY 2629 attcccaaatgagctcttggaagatggttcattcgtgagactacagcgtccttgaca 2688
|||||
Db 1074 ATTCCCAAAATGATCTCTGGGGGATGTTTCATCCGAGAGATTATACAGCGCTTTGACA 1133
QY 2689 gggccaacaacctcgttggagtaagtaagaatctga 2726
|||||
Db 1134 GGGCCAACAACCTCGTGGGGCTGGCCAAAGCATCTGA 1171

Search completed: July 31, 2002, 18:19:33
Job time: 20124 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 21:11:57 ; Search time 790.32 Seconds
(without alignments)
8596.303 Million cell updates/sec

Title: US-09-643-755B-3
Perfect score: 3957
Sequence: 1 ctgcaggaattcatgtact.....acctacaactaagtacc 3957

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_032802: *
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3957	100.0	22	AAS00570
2	1545.2	39.0	20	AAZ10376
3	1539	38.9	20	AAZ10376
4	1226.8	31.0	20	AAZ10392
5	1173	29.6	22	AAS00569
6	1149.8	29.1	22	AAZ17528
7	898.8	22.7	10	AAZ1157
8	897.4	22.7	16	AAZ0306
9	897.2	22.7	5	AAZ40295

10	897.2	22.7	1460	5	AAZ40180	Sequence of recombinant gene. Synt
11	896.4	22.7	1210	12	AAZ014051	2.7 Kbp HindIII fr
12	895.8	22.6	2733	20	AAZ06463	Sequence of prochy
13	895.6	22.6	1278	5	AAZ40055	Pre-prorenin-A ge
14	895.6	22.6	1460	3	AAZ20043	BamHI/SalI insert
15	895.6	22.6	2727	10	AAZ91188	Sequence of prepro
16	894	22.6	1290	4	AAZ30209	CDNA sequence cor
17	890.8	22.5	1314	4	AAZ30049	Sequence of veal c
18	889.8	22.5	1289	4	AAZ30022	Sequence encoding
19	886.8	22.4	1098	11	AAZ04683	Sequence of proren
20	885.2	22.4	1098	4	AAZ30063	BamHI insert from
21	875.4	22.1	2982	10	AAZ91185	Prochymosin (pro
22	866.6	21.9	1175	13	AAZ02949	Gene encoding (pre
23	704.6	17.8	1143	5	AAZ40214	Optimised prochymo
24	686.8	17.4	1104	9	AAZ80001	Human aspartyl pro
25	534	13.5	1140	24	AAZ97159	Prochymosin gene 5
26	382.4	9.7	637	14	AAZ04459	Human stomach cell
27	359.2	9.1	1393	22	AAZ57421	Human full-length
28	359.2	9.1	1409	22	AAZ44608	DNA encoding novel
29	315.4	8.0	1017	23	AAZ79579	Cat pregnancy asso
30	291	7.4	1360	20	AAZ20193	Human stomach cell
31	248.4	6.3	1357	22	AAZ57422	Human gastricsin c
32	248.4	6.3	1366	22	AAZ09486	Human ovarian tumo
33	248.4	6.3	1418	20	AAZ77523	Bovine pregnancy a
34	247.6	6.3	1340	20	AAZ20179	Bovine pregnancy a
35	236.2	6.0	1285	20	AAZ20181	Bovine pregnancy a
36	234.6	5.9	1258	20	AAZ20162	Bovine pregnancy a
37	231.6	5.9	1130	20	AAZ20182	Bovine pregnancy a
38	230.4	5.8	1173	20	AAZ20183	Bovine pregnancy a
39	226.8	5.7	1173	24	AAZ97161	Human aspartyl pro
40	221	5.6	1358	20	AAZ39788	Gastric cancer ass
41	203.8	5.2	1180	23	ABL21921	Drosophila melanog
42	203.8	5.2	3180	23	ABL21920	DNA fragment of pa
43	203	5.1	828	11	AAZ03224	Bovine pregnancy a
44	199.8	5.0	846	11	AAZ03223	
45	197.2	5.0	1168	20	AAZ20171	

ALIGNMENTS

RESULT	ID	Score	Query Match Length	ID	Description
1	AAS00570	3957	100.0	22	AAS00570
2	1545.2	39.0	20	AAZ10376	Bovine phaseolin p
3	1539	38.9	20	AAZ10376	Nucleotide sequenc
4	1226.8	31.0	20	AAZ10392	Phaseolin vulgaris
5	1173	29.6	22	AAS00569	Nucleotide sequenc
6	1149.8	29.1	22	AAZ17528	Bovine pre-pro-chy
7	898.8	22.7	10	AAZ1157	Kunitz soybean try
8	897.4	22.7	16	AAZ0306	Cloned sequence of
9	897.2	22.7	5	AAZ40295	Chymosin open read

PD 01-MAR-2001.
XX
XX 23-AUG-2000; 2000WO-CA00975.
XX
PR 23-AUG-1999; 99US-0378696.
XX
PA (SEMB-) SEMBIOSYS GENETICS INC.
XX
PI Van Rooijen G, Keon RG, Boothe J, Shen Y;
XX
XX WPI: 2001-226621/23.
DR P-PSDB; AAU00536.
XX
XX Producing chymosin in seeds of plants such as rice, flax, rape seed, by
PT transforming plant cell with a nucleic acid encoding chymosin operably
PT linked to transcription regulator and terminator sequences -
XX
XX Example 1; Fig 2; 56pp; English.
XX
XX The sequence represents a chimeric polynucleotide comprising a
CC pre-pro-chymosin, a phaseolin promoter and a phaseolin terminator.
CC Chymosin can be produced in a plant seed through introduction of a
CC chimeric nucleic acid molecule, comprising a nucleic acid sequence
CC encoding a chymosin polypeptide operatively linked to transcription
CC regulator and terminator sequences, into a plant cell. The sequences are
CC useful for producing plant seeds, in particular seeds of soybean, rape
CC seed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats,
CC sorghum, Arabidopsis thaliana, potato, flax/linseed, safflower, oil palm,
CC groundnut, Brazil nut, coconut, castor, coriander, squash, jojoba and
CC rice.
XX
XX Sequence 3957 BP: 1263 A; 790 C; 609 G; 1295 T; 0 other:
SQ
Query Match 100.0%; Score 3957; DB 22; Length 3957;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 3957; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctgcaggaattcatgtactccaglatcatatagtagaagtttggctctcgccg 60
DB 1 ctgcaggaattcatgtactccaglatcatatagtagaagtttggctctcgccg 60
QY 61 tggtttttacctatttaaaagggtttccacctaaattctggtatctctact 120
DB 61 tggtttttacctatttaaaagggtttccacctaaattctggtatctctact 120
QY 121 ttactgttacttaattctcatatacttggttgaaatttcgcgcttcgcgcacga 180
DB 121 ttactgttacttaattctcatatacttggttgaaatttcgcgcttcgcgcacga 180
QY 181 tatccctcaaatattatatttqtaaaacatttcaaacgcgacataaatttatgaagtc 240
DB 181 tatccctcaaatattatatttqtaaaacatttcaaacgcgacataaatttatgaagtc 240
QY 241 ccgtctatcttaattgtagtctaacatttcaatgtgaaatatataatttacttaattt 300
DB 241 ccgtctatcttaattgtagtctaacatttcaatgtgaaatatataatttacttaattt 300
QY 301 agcgttgtagaagaacataaagatttatcttatcttcttcaatgaattgtttaata 360
DB 301 agcgttgtagaagaacataaagatttatcttatcttcttcaatgaattgtttaata 360
QY 361 caataaacaattcttacccttaagaagattcccatattatatttaanaatalat 420
DB 361 caataaacaattcttacccttaagaagattcccatattatatttaanaatalat 420
QY 421 ttatacaaatattttcaaccacgtaaatctcaataaataagttgttcaaaagtaata 480
DB 421 ttatacaaatattttcaaccacgtaaatctcaataaataagttgttcaaaagtaata 480
QY 481 aatttaactccaaattttttatctcgactgacttaagaagaccccgtagacaact 540
DB 481 aatttaactccaaattttttatctcgactgacttaagaagaccccgtagacaact 540

QY 541 agccattttttcttgaataaaaaatccaatatcatctgtattttttatatacaga 600
DB 541 agccattttttcttgaataaaaaatccaatatcatctgtattttttatatacaga 600
QY 601 aaattcaccaacaatcatcttgytatttctcgaagcaagtcagttaagcaattct 660
DB 601 aaattcaccaacaatcatcttgytatttctcgaagcaagtcagttaagcaattct 660
QY 661 ataattccatttgacactacggaagtaacggaagatctgctttacatgagagacat 720
DB 661 ataattccatttgacactacggaagtaacggaagatctgctttacatgagagacat 720
QY 721 ctctcaagtaattttaataagtagtactatataattcaagatttcabatataactcaa 780
DB 721 ctctcaagtaattttaataagtagtactatataattcaagatttcabatataactcaa 780
QY 781 tattctcttaaaaaaattaatagatataataataattacttttttaatttaagtt 840
DB 781 tattctcttaaaaaaattaatagatataataataattacttttttaatttaagtt 840
QY 841 aatttgaatttggactattgatttatttctactatgtttaattgtttatagat 900
DB 841 aatttgaatttggactattgatttatttctactatgtttaattgtttatagat 900
QY 901 agtttaagtaataataagtaatgtagtagagtgtagagtgtagagtgtagagtgtagag 960
DB 901 agtttaagtaataataagtaatgtagtagagtgtagagtgtagagtgtagagtgtagag 960
QY 961 tataacatttaagtgtagactaatttctatataatttctattgtctttacotttttgt 1020
DB 961 tataacatttaagtgtagactaatttctatataatttctattgtctttacotttttgt 1020
QY 1021 atgtgaagtcggaactagaattacagtggttgcacatggaactcgtgtctttgtctc 1080
DB 1021 atgtgaagtcggaactagaattacagtggttgcacatggaactcgtgtctttgtctc 1080
QY 1081 atgcattgggtcttgcgcaagaanaagacaagaanaagaaagaaagacaagagaga 1140
DB 1081 atgcattgggtcttgcgcaagaanaagacaagaanaagaaagaaagacaagagaga 1140
QY 1141 caaaacgcacatcaccaac 1200
DB 1141 caaaacgcacatcaccaac 1200
QY 1201 atgtatgtcttaaatgccatgcaagcaacagtggttcaacatgcaacttaaatgtctac 1260
DB 1201 atgtatgtcttaaatgccatgcaagcaacagtggttcaacatgcaacttaaatgtctac 1260
QY 1261 ccactcacaaccac 1320
DB 1261 ccactcacaaccac 1320
QY 1321 atatatattctctcttcgcgcaccccaatttcttcaacttcaacacacacacacacac 1380
DB 1321 atatatattctctcttcgcgcaccccaatttcttcaacttcaacacacacacacacac 1380
QY 1381 tgggtgtacatcccatgcccacaaatctccatgcatggttccaaacacacttctctataaa 1440
DB 1381 tgggtgtacatcccatgcccacaaatctccatgcatggttccaaacacacttctctataaa 1440
QY 1441 tacctataaatccctataatcatctcaacttcttcaatcatcatcatcatcatcatcat 1500
DB 1441 tacctataaatccctataatcatctcaacttcttcaatcatcatcatcatcatcatcat 1500
QY 1501 ctactctactataataatcccccaacccaacccaacccaacccaacccaacccaacccaac 1560
DB 1501 ctactctactataataatcccccaacccaacccaacccaacccaacccaacccaacccaac 1560
QY 1561 tcccttaagcttcccttctcagcttcccttcttcttcttcttcttcttcttcttcttctt 1620
DB 1561 tcccttaagcttcccttctcagcttcccttcttcttcttcttcttcttcttcttcttctt 1620

OY	1621	ctcagctgctgagatcaccgcgcatctcctctctaacaaagtgatgctctctcgtaagcgc	1680
Db	1621	ctccagctgctgagatcaccgcgcatctcctctctaacaaagtgatgctctctcgtaagcgc	1680
OY	1681	tgaaggaacatggaactctctgaagaactctcttgagaagaacaacagtatgcatcagaacga	1740
Db	1681	tgaaggaacatggaactctctgaagaactctcttgagaagaacaacagtatgcatcagaacga	1740
OY	1741	agtaaccggtctctggtgtaagttgctctagcgtgctgcaacttccaaacttacttbatagtcac	1800
Db	1741	agtaactcggctcttcggtgtaagttgctctagcgtgctgcaacttccaaacttacttbatagtcac	1800
OY	1801	actttgggaagatctacctcccggaaccccgcgctccaagaagttcagcttctcttgatctg	1860
Db	1801	actttgggaagatctacctcccggaaccccgcgctccaagaagttcagcttctcttgatctg	1860
OY	1861	gttcctcttgactcttggtgttcctctctatcttactgccaaggaatgctcctgccaagaaccc	1920
Db	1861	gttcctcttgactcttggtgttcctctctatcttactgccaaggaatgctcctgccaagaaccc	1920
OY	1921	aaaagattcgaatcccgagaagaatctgccaacttccagaagaacttaaggacaacccctgtctat	1980
Db	1921	aaagattcgaatcccgagaagaatctgccaacttccagaagaacttaaggacaacccctgtctat	1980
OY	1981	actacggtacaggtatgacatgacaagaactcttaggtctatgatacgcgtgcactgtctcca	2040
Db	1981	actacggtacaggtatgacatcacaagaactcttaggtctatgatacgcgtgcactgtctcca	2040
OY	2041	ttggggacatttcaacagacggtagagacttagaaccccaagaacaaagtgatgtcttccact	2100
Db	2041	ttggggacatttcaacagacggtagagacttagaaccccaagaacaaagtgatgtcttccact	2100
OY	2101	atgacagaattcgaatgagatctctgtgatagtgatacccaactcgcgtccgtaagtatctg	2160
Db	2101	atgacagaattcgaatgagatctctgtgatagtgatacccaactcgcgtccgtaagtatctg	2160
OY	2161	taactgtgtttgacaacacatgataaacccgaacacttagctcccaagaactgtctcgtgtt	2220
Db	2161	taactgtgtttgacaacacatgataaacccgaacacttagctcccaagaactgtctcgtgtt	2220
OY	2221	acatgggcaaggaatgggccaaggaagagatgtctaagctttggagctttgatccatctact	2280
Db	2221	acatgggcaaggaatgggccaaggaagagatgtctaagctttggagctttgatccatctact	2280
OY	2281	acacagaagatctcttcaactctgtggttcacagtaactgtgagaagatgacggcaattcaactgtg	2340
Db	2281	acacagaagatctcttcaactctgtggttcacagtaactgtgagaagatgacggcaattcaactgtg	2340
OY	2341	acagtgtaaccatcaacgaacggtgtgtgtgtgtgcatgtgaaagtggatgtcaagctatctgtg	2400
Db	2341	acagtgtaaccatcaacgaacggtgtgtgtgtgtgcatgtgaaagtggatgtcaagctatctgtg	2400
OY	2401	ataccggtacgtccaagctgtgtctggacacctagacagcgacacttccaacatccaagcgt	2460
Db	2401	ataccggtacgtccaagctgtgtctggacacctagacagcgacacttccaacatccaagcgt	2460
OY	2461	tttgagaccacaacagaacacagtaacggtgttgatgttgacaatgtgcgacaacacttaagctaca	2520
Db	2461	tttgagaccacaacagaacacagtaacggtgttgatgttgacaatgtgcgacaacacttaagctaca	2520
OY	2521	tgactcaaatgttctcttgagataccaagcgaagaatgtacccaacttgacccctccgctata	2580
Db	2521	tgactcaaatgttctcttgagataccaagcgaagaatgtacccaacttgacccctccgctata	2580
OY	2581	ccagaccagatccaaggttctcttgacaagtgatgttccagaagtgaagacaatccccaagaat	2640
Db	2581	ccagaccagatccaaggttctcttgacaagtgatgttccagaagtgaagacaatccccaagaat	2640
OY	2641	ggactctggagagatgtttatctatcctcgatgataccaagcgtcttcttgaaagggccaacaac	2700
Db	2641	ggactctggagagatgtttatctatcctcgatgataccaagcgtcttcttgaaagggccaacaac	2700
OY	2701	tcgctgttggtctagatcaaggaacatctgaaagcttaataagtaatgaaactaaatgcatgtagt	2760

[illegible]

Db 3781 cactacataaaccttttagcagtagagcaatggtgaccgtgtgttagcttcttla 3840
Qy 3841 tttaattttttatcacgaagaataataataataatgagacactcaagatgtttc 3900
|||||
Db 3841 tttaattttttatcacgaagaataataataataatgagacactcaagatgtttc 3900
Qy 3901 aacctatatacaaaaccccaaaacaagtttccttagcaccctaccaactaagttacc 3957
|||||
Db 3901 aacctatatacaaaaccccaaaacaagtttccttagcaccctaccaactaagttacc 3957
RESULT 2
AAZ10376
ID AAZ10376 standard; DNA: 1558 BP.
XX
AC AAZ10376;
XX
DT 15-NOV-1999 (first entry)
XX
DE Nucleotide sequence of the bean phaseolin promoter.
XX
KW Fao gene: plant metabolism; fatty acid oxidation enzyme;
KW polyhydroxyalkanoate; oil composition; seed production; plant biomass;
KW transgenic plant; promoter; bean; ss.
XX
OS Phaseolus sp.
XX
PN MO9945122-A1.
XX
PD 10-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US04999.
XX
PR 06-MAR-1998; 98US-0077107.
XX
PA (META-) METABOLIX INC.
XX
PI Boynton L, Huismann GW, Moloney M, Patterson N, Peoples OP;
PI Snell K;
XX
DR WPI: 1999-540850/45.
XX
PT Modifying fatty acid metabolism in plants, useful for increasing
PT biomass and producing specific polymers in seeds
XX
PS Example 3: Page 63; 79pp; English.
XX
CC The present sequence represents the bean phaseolin promoter, which
CC is used to construct plasmids for the expression of the P. putida
CC facAB gene. The gene encodes an enzyme that may be used in the
CC method of the invention. The specification describes a method for
CC manipulating the metabolism of a plant, and comprises expressing a
CC heterologous gene encoding fatty acid oxidation enzymes in the
CC cytosol or plastids other than the peroxisomes, glyoxisomes or
CC mitochondria of the plant. The method may be used to enhance the
CC biological production of polyhydroxyalkanoates or novel oil compositions
CC in a transgenic plant. Plants which may be used to produce these
CC compounds in this way include Brassicas, maize, soybean, cottonseed,
CC sunflower, palm, coconut, safflower, peanut, mustard, flax, tobacco and
CC alfalfa. The method may also be used to prevent or suppress seed
CC production and therefore increase the production of biomass (leaves,
CC stems, stalks) by plants.
CC
SQ Sequence 1558 BP; 516 A; 310 C; 171 G; 561 T; 0 other;

Qy 66 tttaaccttaatttaagggttttccacctaataaaitcgtgatacattctacttact 125
|||||
Db 62 tttaaccttaatttaagggttttccacctaataaaitcgtgatacattctacttact 121
Qy 126 tgttaacttaatttccataaactcttggttggaattatcaagcttcgcgcagatattcc 185
Db 122 tgttaacttaatttccataaactcttggttggaattatcaagcttcgcgcagatattcc 181
Qy 186 ctacaatttattttgtttaacattttcaaacccgataaattttatgagcccgctc 245
Db 182 ctacaatttattttgtttaacattttcaaacccgataaattttatgagcccgctc 241
Qy 246 tatcttaatgtagtctaacaatttccatbattgaaatataataattacttaattgaagc 305
Db 242 tatcttaatgtagtctaacaatttccatbattgaaatataataattacttaattgaagc 301
Qy 306 tggtagaagaagataaagattatcttattcttcttccataaagtgtttaataataa 365
Db 302 tggtagaagaagataaagattatcttattcttcttccataaagtgtttaataataa 361
Qy 366 taacaacatttcttaaccttaagaagattcccatatttataatttaaaatataattatc 425
Db 362 taacaacatttcttaaccttaagaagattcccatatttataatttaaaatataattatc 421
Qy 426 aaataatttcaaccacgtaaatctcataaataaagttgttcaaaaagtaataaattt 485
Db 422 aaataatttcaaccacgtaaatctcataaataaagttgttcaaaaagtaataaattt 481
Qy 486 aacctcataattttttatctgactgtatcttaaggacacacccagtgacaaactagcca 545
Db 482 aacctcataattttttatctgactgtatcttaaggacacacccagtgacaaactagcca 541
Qy 546 ttttttcttggataaataaataatccaattatcatgttattttttatatacaatgaaatt 605
Db 542 ttttttcttggataaataaataatccaattatcatgttattttttatatacaatgaaatt 601
Qy 606 tcaaccaacaatcatgttggatattctgaaagcaagtcagttagtgcataatctcataat 665
Db 602 tcaaccaacaatcatgttggatattctgaaagcaagtcagttagtgcataatctcataat 661
Qy 666 tccatttggacactacggaagtaactggaagctcgttttccatgagcgagacacatctct 725
Db 662 tccatttggacactacggaagtaactggaagctcgttttccatgagcgagacacatctct 721
Qy 726 aaagtaatttaataatagttactatatactcaagatttccatatataatcaataatata 785
Db 722 aaagtaatttcaataatagttactatatactcaagatttccatatataatcaataatata 781
Qy 786 ctcttaaaaatlaattagataataataaataattacttttlaatttaagttaattg 845
Db 782 ctcttaaaaatlaattagataataataaataattacttttlaatttaagttaattg 841
Qy 846 ttgaatttggactcttattattattattctctctctgttgaatttggtttagatagtt 905
Db 842 ttgaatttggactcttattattattattctctctctgttgaatttggtttagatagtt 901
Qy 906 aaagtaaatataaagtaagtagtagagtgtagagtgtagccttaaacataaactataa 965
Db 902 aaagtaaatataaagtaagtagtagagtgtagagtgtagccttaaacataaactataa 961
Qy 966 cattatgttgactaatttccatatattcttattgtccttactccttctgtgtagta 1025
Db 962 cattatgttgactaatttccatatattcttattgtccttactccttctgtgtagta 1021
Qy 1026 agtcgtaactagaattcagtggttgccatgagactctgtgcttttgtagta 1085
Db 1022 agtcgtaactagaattcagtggttgccatgagactctgtgcttttgtagta 1081
Qy 1086 tgggtcttgcgcgaagaaagacaaagacaaagaaagacaaagacagagagacaa 1145
Db 1082 tgggtcttgcgcgaagaaagacaaagacaaagaaagacaaagacagagagacaa 1141

QY	1146	cgcaatcacacaacccaactctaattgtgctacgtgctgatacgaatcgcgcggttcagtta	1205
Db	1142	cgcatacacacaacccaactctaattgtgctacgtgctgatacgaatcgcgcggttcagtta	1201
QY	1206	tgcttaataatgcataatgcaagcaacacagtgcttaacatgcaacttaaatggctccaccac	1265
Db	1202	tgcttaataatgcataatgcaagcaacacagtgcttaacatgcaacttaaatggctccaccac	1261
QY	1266	tcaaccacacacacaacacacatgtgcctttttcttcatacatacacaacacacgtgataa	1325
Db	1262	tcaaccacacacacaacacacatgtgcctttttcttcatacatacacaacacacgtgataa	1321
QY	1326	tctacttctcttcgcgcacactccaattttcttccaacttcaacacacgtaaacctgcatatgct	1385
Db	1322	tctacttctcttcgcgcacactccaattttcttccaacttcaacacacgtaaacctgcatatgct	1381
QY	1386	gtcatcccatgcccacaattctcgaatgtatgttccaacacctctctctataataact	1445
Db	1382	gtcatcccatgcccacaattctcgaatgtatgttccaacacctctctctataataact	1441
QY	1446	ataaataaccttaataatcactactcttcttcatacgcataccatccagagtaactact	1505
Db	1442	ataaataaccttaataatcactactcttcttcatacgcataccatccagagtaactact	1501
QY	1506	ctactataataataatcccccaacccaactcatatcatactactactactat	1555
Db	1502	ctactataataataatcccccaacccaactcatatcatactactactactactat	1551

RESULT	3
AAx90961	
ID	AAx90961 standard; DNA; 1547 BP.

AC AAX90961;

DT 17-JAN-2000 (first entry)

DE Phaseolus vulgaris beta phaseolin gene 5' regulatory sequence.

KM palmitate-CoA delta-9 desaturase; PCA; beta phaseolin gene; fatty acid;
 KM palmitic acid; desaturate; modification; saturate oil; transgenic plant;
 KM 5' regulatory element; plasmid pGm184-2; tobacco; transformed plant; ss

OS *Phaseolus vulgaris*.

...	key	Location/Qualifiers
FH	old_sequence	replace(1049, C)
FT		

FTT /note= "Corresponds to base 1113 of plasmid phagN184-2
FTT and facilitates subsequent cloning"

PN W09950430-A2.

PD 07-OCT-1999.

PF 29-MAR-1999; 99WO-US06765.

PR 30-MAR-1998; 98US-0079840.

PA (DOWC) DOW AGROSCIENCES LLC.

PI Folkerts O, Merlo DJ;

DR WPI; 1999-610860/52.

PT New isolated

XX

XX

CC gene from Ph

cc plasmid phag

CC	and seed specific expression in tobacco. This plasmid can be used for
CC	producing transgenic plants having altered saturated oil profiles and
CC	decreased fatty acid levels in transformed plants.
xx	
SQ	Sequence 1547 BP; 513 A; 306 C; 170 G; 558 T; 0 other:
	Query Match 38.9%; Score 1539; DB 20; Length 1547;
	Best Local Similarity 99.7%; Pred. No. 1.4e-208;
	Matches 1542; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY	6 ggaatcatctgtaccccgatcattatagtgaaagtcttgcctctccgcytgtt 65
DB	1 gaattccattgtacctcccgatacatatagtgaaagtttcgtcctccgcygtgt 60
QY	66 tttaacctattaagaagggtcttcaccctaaaattctgttcaatctcaattact 125
DB	61 tttacccttatttaaagggttctccaccctaaaattctgttcaatctcaattact 120
QY	126 tgttaacttaatttcacaalccttgyttgtaathatalocagcttcgcacagalcc 185
DB	121 tgttaacttaatttcacataactcttgyttgaathatalocagcttcgcacagalcc 180
QY	186 ctacaaatttatatttgyttaaacatttccaacgcacaaaaatttatgaagtcgcyc 245
DB	181 ctcaaatattatatattgttaaacaatttccaacgcacaaaaatttatgaagtcgcyc 240
QY	246 tatctttaatgtagtctaaccatttccatattgaalatalaatttacttaatttagcgt 305
DB	241 tatctttaatgtagtctaaccatttccatattgaalatalaatttacttaatttagcgt 300
QY	306 tggtagaagcacaaagattatctctatctcttcacataaagtttaatacaata 365
DB	301 tggtagaagcacaaagattatctctatctcttcacataaagtttaatacaata 360
QY	366 taacaaactcttccactaagaaggattcccatttatattttaaaaaatattatc 425
DB	361 taacaaactcttccactaagaaggattcccatttatattttaaaaaatattatc 420
QY	426 aaatatttttcaacacagfnaatlcocatnaatnatagttggttcaaagaataaaatt 485
DB	421 aaatatttttcaacacagfnaatlcocatnaatnatagttggttcaaagaataaaatt 480
QY	486 aactcataattttttatcgcagcatccttaagaacaacccagtgacacaactagcca 545
DB	481 aactcataattttttatcgcagcatccttaagaacaacccagtgacacaactagcca 540
QY	546 ttttttcttggataaaaaaalccaatatcatgtfatTTTTTTTatacaatgaaatt 605
DB	541 ttttttcttggataaaaaaalccaatatcatgtfatTTTTTTTatacaatgaaatt 600
QY	606 ttaaccaacaatcatttgygfatcttcggaagaagcatgltatgacaaatctcataat 665
DB	601 ttaaccaacaatcatttgygfatcttcggaagaagcatgltatgacaaatctcataat 660
QY	666 tcccatittgacactgcggaagtaactggaagctgcttttacatgcgagacacatctct 725
DB	661 tcccatittgacactgcggaagtaactggaagctgcttttaacatgcgagacacatctct 720
QY	726 aaagtaattttaataatagttiaclataitccaagattcaatalacaaatactcaatata 785
DB	721 aaagtaattttaataatagttiaclataitccaagattcaatalacaaatactcaatata 780
QY	786 cttctraaaaanaatattgatgataataatnaaatattactcttttaattttaagttaattg 845
DB	781 cttctraaaaanaatattgatgataataatnaaatattactcttttaattttaagttaattg 840
QY	846 ttgaatttgtagactatgatattatattctactactggtttaaatggttttatagatagttt 905
DB	841 ttgaatttgtagactatgatattatattctactactggtttaaatggttttatagatagttt 900
QY	906 aaagtaaatataagtaaatgtagtagagtggttagaggttttaacccttaaacataactata 965

—

Dg	901	aaagtaactaagtatgtatgtatgaagtgtagatgtagtgaagtgtagtacccttaaaccaataactata	960
Qy	966	cattatggtgagcctaattttcataatbttcttattgtcttcttacccttctcttgtagta	1025
Db	961	gattatggtgagcctaattttcataatbttcttattgtcttcttacccttctcttgtagta	1020
Qy	1026	agtcgcgaactaataattacaaatggtgtggtgtgcgaagacctgtgtgtctttgtgtcatgca	1085
Db	1021	agtcgcgaactaataattacaaatggt	1080
Qy	1086	tgggtcttcgcaagaagaaagacaaagaaacaaagaaagaaagaaagaaagaaagaaagaa	1145
Db	1081	tgggtcttcgcaagaagaaagacaaagaaacaaagaaagaaagaaagaaagaaagaaagaa	1140
Qy	1146	cgcgaatcaccaaccac	1205
Db	1141	cgcgaatcaccaaccac	1200
Qy	1206	tgttaaatgccatgcaaaagcaaacagctgtcttaacatgcaactttaatgtgtcacacac	1265
Db	1201	tgttaaatgcatgcaaaagcaaacagctgtcttaacatgcaactttaatgtgtcacacac	1260
Qy	1266	tcaaccac	1325
Db	1261	tcaaccac	1320
Qy	1326	ttcatctcttcgcgcac	1385
Db	1321	ttcatctcttcgcgcac	1380
Qy	1386	gtcatcccatgcccacaaatctcaatgcacatgtctccaaacacacttctcttataataact	1445
Db	1381	gtcatcccatgcccacaaatctcaatgcacatgtctccaaacacacttctcttataataact	1440
Qy	1446	ataaatcccttaatatcatcatcactctcttcatcatcaccatccagagtaactact	1505
Db	1441	ataaatcccttaatatcatcatcactctcttcatcatcaccatccagagtaactact	1500
Qy	1506	ctactactataatacccccaaccccaactcatatcaataactactac	1552
Db	1501	ctactactataatacccccaaccccaactcatatcaataactactac	1547
RESULT 4			
AAZ10392			
ID	AAZ10392	standard; DNA; 1244 BP.	
XX	AC	AAZ10392;	
XX	DT	15-NOV-1999 (first entry)	
DE	Nucleotide sequence of the bean phaseolin terminator.		
XX			
KW	Fao gene; plant metabolism; fatty acid oxidation enzyme;		
KW	polyhydroxyalkanoate; oil composition; seed production; plant biomass;		
KW	transgenic plant; terminator; bean, ss.		
XX			
OS	Phaseolus sp.		
XX			
PN	W09945122-A1.		
XX			
PD	10-SEP-1999.		
XX			
PF	05-MAR-1999; 99WO-US04999.		
XX			
PR	06-MAR-1998; 98US-0077107.		
XX			
PA	(META-) METABOLIX INC.		
XX			
PI	Boynton L, Huismann GW, MoJoney M, Patterson N, Peoples OP;		
XX	Snell K;		
DR	WPI; 1999-540850/45.		

XX Modifying fatty acid metabolism in plants, useful for increasing
PT biomass and producing specific polymers in seeds
XX
XX
PS Example 3: Page 71; 79pp; English.
XX
CC The present sequence represents the bean phaseolin terminator, which
CC is used to construct plasmids for the expression of the P. putida
CC faabA gene. The gene encodes an enzyme that may be used in the
CC method of the invention. The specification describes a method for
CC manipulating the metabolism of a plant, and comprises expressing a
CC heterologous gene encoding fatty acid oxidation enzymes in the
CC cytosol or plastids other than the peroxisomes, glyoxisomes or
CC mitochondria of the plant. The method may be used to enhance the
CC biological production of polyhydroxyalkanoates or novel oil compositions
CC in a transgenic plant. Plants which may be used to produce these
CC compounds in this way include Brassicas, maize, soybean, cottonseed,
CC sunflower, palm, coconut, safflower, peanut, mustards, flax, tobacco and
CC alfalfa. The method may also be used to prevent or suppress seed
CC production and therefore increase the production of biomass (leaves,
CC stems, stalks) by plants.
XX
SQ Sequence 1244 BP; 456 A; 174 C; 178 G; 436 T; 0 other;

Query Match	31.0%	Score 1226.8	DB 20	Length 1244	
Best Local Similarity	99.4%	Pred. No. 1.4e-164			
Matches 1231:	Conservative	0	Mismatches 7	Indels 0	Gaps 0
QY	2720	aatcgaagcttaataaagatgaactaaatgcagtatggtgtaagagctcattgagagc	2779		
Db	7	aatgcgaagcttaataaagatgaactaaatgcagtatggtgtaagagctcattgagagc	66		
QY	2780	atggaatatgtatccgacacatgtaacgtaataaacctgagctccatctcctctc	2839		
Db	67	atggaatatgtatccgacacatgtaacgtaataaacgagctccatctcctcctc	126		
QY	2840	atggaataacaaagagatgtaatgataataaactcctcactcattgacctatgttcta	2899		
Db	127	atggaataacaaagagatgtaatgataataaactcctcactcattgacctatgttcta	186		
QY	2900	tgaataattccctcttatattataaatacatctgataatcgtagagcgcttatggatgcttc	2959		
Db	187	tgaataattccctcttatattataaatacatctgataatcgtagagcgcttatggatgcttc	246		
QY	2960	aaatagtacaaaacaaatggtgactatgaagctttcgaacaattcttaacttagcatt	3019		
Db	247	aaatagtacaaaacaaatggtgactatgaagctttcgaacaattcttaacttagcatt	306		
QY	3020	gtgaacgagacataaagtgttlaagaagacataacaatttaatggaagaagtttgtctcca	3079		
Db	307	gtgaacgagacataaagtgttlaagaagacataacaatttaatggaagaagtttgtctcca	366		
QY	3080	tttatatattatataattaccocattatgtaattatataatgagatgtaagaagacataacaa	3139		
Db	367	tttatatatattatataattaccocattatgtaattatataatgagatgtaagaagacataacaa	426		
QY	3140	ttaataaaggagaagttttatccattatattatattatattatattacacccattatataattata	3199		
Db	427	ttaataaaggagaagttttatccattatattatattatattatattatattacacccattatataattata	486		
QY	3200	cttatccactattttaatgctcttataaagtttgtaglccatgabatatttctaattttagtt	3259		
Db	487	cttatccactattttaatgctcttataaagtttgtaglccatgabatatttctaattttagtt	546		
QY	3260	tgaatgatatgaaagggtactatttggaactccttactcgtataaagggttgagctat	3319		
Db	547	tgaatgatatgaaagggtactatttggaactccttactcgtataaagggttgagctat	606		
QY	3320	ccttaagaagggtcctttaatttatgttcttcacagataaaaaaaattatagaatt	3379		
Db	607	ccttaagaagggtcctttaatttatgttcttcacagataaaaaaaattatagaatt	666		

OY		3380	ggtttgatataaatattgaaggattttaaaatataatatataataataacatacata	3439
Dd		667	ggcttgatataaaaataatggaagatttaaaaaataataataataataacaatacaata	726
OY		3440	tgtatataatttattcatataaacatttatcvtataaaaaagaanaatlattytcataat	3499
Dd		727	tgtatataatttattcatataatacaactttatctataaaaaagaanaatlattytcataat	786
OY		3500	cataacaatcgitttagccitgcgtgcgaacctccaattatlttaaocqagytlaaacat	3559
Dd		787	ctatacaatcgitttagccitgcgtgcgaacctccaattatlttaaogagylaaacarat	846
OY		3560	ttgaccttttgtgtattttaacaatatattttaoasctalatyaaatTTTTTTTtTa	3619
Dd		847	ttgaccttttgtgtattttaacaatatattttaoasctalatayaaatTTTTTTTtTa	906
OY		3620	ttggcgaaggaataaaatttaatttagagaggacaatgytgytccccaatcccatacac	3679
Dd		907	ttggcgaaggaataaaatttaatttagagaggacaatgytgytccccaatcccatacac	966
OY		3680	caaattccaccagaagaagtgcaggtccggggacacaacaaaaaacaggcaagggaattttta	3739
Dd		967	caaattccaccagaagaagtgcaggtccggggacacaacaaaaaacaggcaagggaattttta	1026
OY		3740	atttgggttgcttgcttgctgcgcataattaatgcagfnaaaacataacatacccttl	3799
Dd		1027	atttgggttgcttgcttgctgcgcataattaatgcagfnaaaacataacatacccttl	1086
OY		3800	aaccagtagagccaatggttgacgccgtgtgcttagcttcctttatlttttttacagca	3859
Dd		1087	aaccagtagagccaatggttgacgccgtgtgcttagcttcctttatlttttttacagca	1146
OY		3860	aagaataaaataaaataaaatagagacacttaaggatgtttcaacccttatacaacaaaccc	3919
Dd		1147	aagaataaaataaaataaaatagagacacttaaggatgtttcaacccttatacaacaaaccc	1206
OY		3920	aaaaaaaaagtttccttagcacccctcacactaactaagtacc	3957
Dd		1207	aaaaaaaaagtttccttagcacccctcacactaactaagtacc	1244
<hr/>				
RESULT	5			
AAS00569				
ID	AAS00569	standard; DNA;	1173 BP.	
XX	AC	AAS00569;		
XX	DT	14-MAY-2001	(first entry)	
DE	Bovine pre-pro-chymosin DNA sequence.			
XX	KW	Chymosin; transcription regulator; terminator sequence; soybean; corn;		
KW	KM	pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;		
RW	KV	bareley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;		
KW	KW	safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;		
XX	XX	squash; jojoba; ds.		
OS	Bos sp.			
FH	Key	Location/Qualifiers		
FT	CDS	/tag= a		
FT		1..1173		
FT		/product= "Bovine chymosin"		
FT		1..78		
FT		/tag= b		
FT		79..201		
FT		/tag= c		
FT		/note= "Pro sequence"		
FT		202..1170		
FT		/tag= d		
FT		/product= "Mature bovine chymosin"		
NN	M0200114571-AI.			

Query Match	29.6%; Score 1173;	DB 22;	Length 1173;
Best Local Similarity	100.0%;	Pred. No. 5.3e-157;	
Matches 1173;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY 1554 atgaacttcctaaatcttcccttcttaagcttcccttggtttggtcaacttcgtt 1613			
Db 1 atgaacttcctaaatcttcccttcttaagcttcccttggtttggtcaacttcgtt 60			
QY 1614 gctgttaactcaagcgcgtcgagatcacccgcattctctctaaaggaagcttcct 1673			
Db 61 gctgttaactcaagcgcgtcgagatcacccgcattctctctaaaggaagcttcct 120			
QY 1674 aagcgctgaagaaacatgagcttctagaagactctctgcagaacaacagatgacatc 1733			
Db 121 aagcgctgaagaaacatgagcttctagaagactctctgcagaacaacagatgacatc 180			
QY 1734 agcagcaagctactccggttcggfgaagttgtagcggtgcacattaccaacttcgtat 1793			
Db 181 agcagcaagctactccggttcggfgaagttgtagcggtgcacattaccaacttcgtat 240			
QY 1794 agtcaactcttgggaagaatctactctgcgaaccccgctcaagagttacggttccttt 1853			
Db 241 agtcaactcttgggaagaatctactctgcgaaccccgctcaagagttacggttccttt 300			
QY 1854 gatactggttccctcgcgactctcgggtttccctctatctactcgaagaagcaatgcctgcgaag 1913			
Db 301 gatactggttccctcgcgactctcgggtttccctctatctactcgaagaagcaatgcctgcgaag 360			
QY 1914 aaccacaaagatctgcagtcgagaaagtcgtccacacttcacgaacttaggcgaaccccttg 1973			
Db 361 aaccacaaagatctgcagtcgagaaagtcgtccacacttcacgaacttaggcgaaccccttg 420			
QY 1974 tctataacactcaggtgacaggtgacgtggaagaagacttaggtatgataacggttccttc 2033			
Db 421 tctataacactcaggtgacaggtgacgtggaagaagacttaggtatgataacggttccttc 480			
QY 2034 tccaacactcttggtgacatctcaacagcaatagagcttagcacccaagaacaggtatctc 2093			
Db 481 tccaacactcttggtgacatctcaacagcaatagagcttagcacccaagaacaggtatctc 540			

OY	2094	ttacaccatgacgaattcgatggcaccctctgtagtgcataccctgcctgcgcgcagag	2153
Db	541	ttccaccatgacaaatctcgatgcataccctctgtatagtcataccctgcctgcgcgcagag	600
OY	2154	tactcgaataccttgctttgacacaatgaaacgcgaacacctgatagtccaaagtgttc	2213
Db	601	taccgcgataacctgtgtttgacacaatgatgaaacgcgaacctagatgcctcaagtgttc	660
OY	2214	tcgggttacatgacacgaagaatggtccagagagcatgctcagctgtgagctattgatcca	2273
Db	661	tcggtttacatacgtgacaggaatggtccagagagacatgctcagctgtgagctattgatcca	720
OY	2274	tcctactacaacagagatctctctcaactgtgttccagatcaactgtgtcagacgatactgtgcaattc	2333
Db	721	tcctactacaacagagatctctctcaactgtgttccagatcaactgtgtcagacgatactgtgcaattc	780
OY	2334	actgtgacacagtgctcaaccatcacagcggctgtgtgtatgtcatacgtgaaggtgagatgtcaagct	2393
Db	781	actgtgacacagtgctcaaccatcacagcggctgtgtgtatgtcatacgtgaaggtgagatgtcaagct	840
OY	2394	atcttggatatccggttaagtctcaagcgtgtgtccgaacctcagaagcgaaattctcaaatcttag	2453
Db	841	atcttggatatccggttaagtctcaagcgtgtgtccgaacctcagaagcgaaattctcaaatcttag	900
OY	2454	caagcgtattgtaggcacacacgaacacgactcaggtgagattgacatagattgcgcacaacct	2513
Db	901	caagcgtattgtaggcacacacgaacacgactcaggtgagattgacatagattgacgcacaacct	960
OY	2514	agctacatgctctacagtgtgtctttgagatcaacgcgcagaatgtaccacactgacccctcc	2573
Db	961	agctacatgctctacagtgtgtctttgagatcaacgcgcagaatgtaccacactgacccctcc	1020
OY	2574	gctcatccacgcgcaggttaagaggttctgcacacagtgagattccaaagtgtgaagaacattcc	2633
Db	1021	gctcatccacgcgcaggttaagaggttctgcacacagtgagattccaaagtgtgaagaacattcc	1080
OY	2634	cagaatattgatactgtggagagatgtgtttcaattcgtgagatactaaagcgtcttttacagggcc	2693
Db	1081	cagaatattgatactgtggagagatgtgtttcaattcgtgagatactaaagcgtcttttacagggcc	1140
OY	2694	aacaacctcgttgtgcctagctaaagcaatctga	2726
Db	1141	aacaacctcgttgtgcctagctaaagcaatctga	1173

XX	RESULT	6
XX	AAD17528	
XX	ID	AAD17528 standard; DNA; 2970 BP.
XX	AC	AAD17528;
XX	DI	10-DEC-2001 (first entry)
XX	DE	Kunitz soybean trypsin inhibitor (KSTI) encoding DNA.
XX	KM	Soybean: allergen: transgenic plant; P34 protein; Gly m Bd 30K; Gly m 1.
XX	KW	alabId; food; infant formula; animal feed; coating; salad oil; syrup;
XX	KW	spraying oil; roasting oil; frying oil; cracker; confectionery product;
XX	KX	snack food; topping; sauce; batter; breading mixture; baking mix; dough;
XX	KX	Kunitz soybean trypsin inhibitor; KSTI; KTI3; ds.
OS	Glycine max.	
XX	PN	WO200168887-A2.
XX	PD	20-SEP-2001.
XX	PF	15-MAR-2001; 2001WO-US08254.
XX	PR	16-MAR-2000; 2000US-0189823.
XX		

PA (DUPO.) DU PONT DE NEMOURS & CO E. I.
PA (PLON-) PIONEER HI-BRED INT INC.
XX
XX
XX Jung R, Kinney AJ;
PI WPI: 2001-582460/65.
DR
XX Recombinant expression construct to lower allergen (e.g., Gly m Bd 30K)
PT content of a soybean, comprises a nucleic acid fragment encoding the
PT allergen, useful for producing soybean plants which can be used to make
PT soybean products -
XX
XX
PS Claim 4; Page 48-49; 57pp; English.
XX
CC The patent discloses hypoallergenic transgenic soybeans and recombinant
CC expression constructs to lower soybean vacuolar protein, commonly known
CC as P34 (Gly m Bd 30K or Gly m 1) and other allergens such as Gly m IA,
CC Gly m IB, rGly m3 or Glycinin G1 (IaabiB). The allergen content of the
CC soybean is reduced by sense suppression which is accomplished by using
CC the expression construct that comprises a nucleic acid fragment encoding
CC the allergen. The constructs are useful for producing hypoallergenic
CC transgenic soybean plants which can be used to make hypoallergenic
CC soybean products which can be used in a variety of food (e.g., infant
CC formulas) and animal feed applications. The oil made from seeds of the
CC hypoallergenic transgenic soybean plants can be used as ingredients,
CC as coatings, as salad oils, as spraying oils, as roasting oils, and
CC as frying oils. The foods in which the oil may be used include crackers
CC and snack foods, confectionery products, syrups and toppings, sauces,
CC batter and breaded mixtures, baking mixes and doughs. The present
CC sequence is a DNA encoding knifit soybean trypsin inhibitor (KSTI
CC or KTI3), a minor soybean seed allergen.

Sequence 2970 BP; 1018 A; 547 C; 527 G; 878 T; 0 other;

[illegible]

Db	669	accgcctcactacaacagggtccctccactcgtggtgcccgtgacagtgacaglaactgac	728
Qy	2329	aattcaactgttgacaagttgacaccatcagcagcgtgtggtgtgtgtgacatgtaaggtgac	2388
Db	729	agttcaactgttgacagtggtcaacatcagcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt	788
Qy	2389	aagatcatcttggataccggtacgtgtccaaagctgtgtcggacctagcagcagacatctcaaca	2448
Db	789	aggcatctcttgacaacggcagccctccaagctgtgtcggccagcagcagacatctcaca	848
Qy	2449	ttcacaacagctatttggaaagccacacagaaacccagtaacggtgagttgacatagattggaca	2508
Db	849	tccagcagggccatttggaaagccacacagaaacccagtaacggtgagttgacatagattggaca	908
Qy	2509	accttagctacatgctcactacagttgtctcttgatatacagcggcagaagatttacccaatgacc	2568
Db	909	accttagctacatgctcactacagttgtctcttgatatacagcggcagaagatttacccaatgacc	968
Qy	2569	ccttcgcctatcacagcaggaatcaagggtctctgacacagtgatctcagaagtgaagacc	2628
Db	969	ccttcgcctatcacagcaggaagcaggaagcgtctctgtaccagtggtcttcagagtgaaatc	1028
Qy	2629	attcccaagaatlgatcttgggaagtgtttcatctgtgtgactaagcgtctttgaca	2688
Db	1029	attcccaagaatlgatcttgggaagtgtttcatctgtgtgactaagcgtctttgaca	1088
Qy	2689	gggccacaacacctcgttggcgtagctaaagcaatctga	2726
Db	1089	gggccacaacacctcgttggcgtagctaaagcaatctga	1126
RESULT 8			
AAT03006			
ID	AAT03006	standard; DNA; 1240 BP.	
AC	AAT03006;		
XX			
DT	13-JUN-1996	(first entry)	
XX			
DE	Chymosin open reading frame.		
XX			
KM	a1-3; albino mutant; light-regulated; Neurospora; bread mould;		
KW	heterologous gene; expression; control; chymosin; ss.		
XX			
OS	Mammalian sp.		
PN	W09530739-A1.		
XX			
PD	16-NOV-1995.		
XX			
PF	09-MAY-1995; 95MO-US05716.		
XX			
PR	10-MAY-1994; 94US-0240372.		
XX			
PA	(UYHA-) UNIV HAWAII.		
XX			
PI	Kato EK, Stuart WD;		
XX			
DR	WPI; 1995-404108/51.		
XX			
PT	Nucleic acid for expression of heterologous protein - contains		
XX	albino promoter for light induced expression in filamentous fungi		
XX			
PS	Example 2; Fig 7; 26pp; English.		
XX			
CC	The DNA is that of a mammalian gene (open reading frame) encoding		
CC	chymosin. The gene was placed in operable linkage with the a1-3		
CC	promoter (see AAT03005). The a1-3 gene controls the production of		
CC	geranyl geranyl pyrophosphatase (GGPP) synthase. GGPP is a precursor		
CC	for carotenoids and xanthophylls. It has been shown that exposure to		
CC	light increases the transcription level of GGPP synthetase 15-45 fold.		
CC	Light activates a number of genes in the common bread mould, Neurospora.		
CC	This can be used to regulate the expression of genes encoding		

CC heterologous proteins, e.g. chymosin, in recombinant production systems.
CC use of a light-regulated promoter is a simple and effective way to
CC control expression and allows timing to be adapted to the physiological
CC status of the host.
XX
SQ Sequence 1240 BP; 274 A; 374 C; 339 G; 253 T; 0 other;

Query Match	22.7%	Score 897.4	DB 16	Length 1240
Best Local Similarity	87.4%	Pred. NO. 3.6e-118		
Matches 982; Conservative	0	Mismatches 141	Indels 0	Gaps 0

QY	1609	tcgtctgctgttactcaacgcctgcgtcgagataaccgcgatcttcctctctcaaaagttaagttctc	1668
Db	100	ttcttcgcctctctcccaaggcgtcgtgaataaccacggagatcttcctctgttacaaggcaagttctc	159
QY	1669	tcgcgttaagcgtctbaaaggaaacatcgagctctcttaagaactctcttgagaaacaacagtatg	1728
Db	160	tcgagaaagcgcctgaaagagacagaggcctctctgtaagatcttccttcgacagaaacagcagtatg	219
QY	1729	gcacacgacgaagaatctaccggctctcgttgaaattctctagacgtgcgcacttaaccaactacc	1788
Db	220	gcacacgacgaagaatctaccggctctcgttgaaattctctagacgtgcgcacttaaccaactacc	279
QY	1789	ttgatagtacaatattcttggaagaactactaccctcgaaaccccgccctcaagaagttcacggttc	1848
Db	280	ttgatagtacagtaattcttggaagaactactaccctcgaaaccccgcccaagaagtctcacggttc	339
QY	1849	tccttgatactggttcctctcgacttctcgttccctctctactacgcaagaagaaatgctc	1908
Db	340	tgcttgaaactggtcctctcgacttctcgttccctctctactacgcaagaagaaatgctc	399
QY	1909	gcaagaaccacccaagaatctcgatccgagaagtcgtccacctctccagaacttgaacaac	1968
Db	400	gcaaaaaacacccgcgtctcgacccggaagaatgcgtccacctctccagaactcctgggcaacg	459
QY	1969	ccttgctatacactacacgtgatacagtgatgacatgcaagaagaacttaagctatagtataccgtca	2028
Db	460	ccctgctataccactacaggaacagcagcatggaaggacatctctagctatagtataccgtca	519
QY	2029	ctgctccoaacattgtgagacattccaacagacgttagacttagacacccaagaacagagtg	2088
Db	520	ctgctccoaacattgtgagacatccagagagacagtaagcccttgacacccaagaagcccgagg	579
QY	2089	atgctctcaacctgaacagaaatctcgatgagacatcccttgatattggacatacccatctgcctggt	2148
Db	580	acgctctcaacctgaacagaaatctcgagatcccttgagatggctcaccctctgcctgcctc	639
QY	2149	cagagtaactcgatacctgtgtctgacaacaatgaatgaaccggaacactagtagtcaagaact	2208
Db	640	cagagtaactcgatacccggtgtgtgacaacaatgaatgaagaagcactgtggtggcccaagacc	699
QY	2209	tgcttcctcggtttacaatgacaaggaatggtccaggaagagcatgtcatcgccttgagactatg	2268
Db	700	tgcttcctcggtttacaatgacaaggaatggtccaggaagagcatgtcatcgccttgagactatg	759
QY	2269	atccatccctaactacaacggatctctctcaactcgtgttcacataactgtgcagcagtatcgtgc	2328
Db	760	accgcgtctactacaacagggttccctcgactggtgtgcctgtgaacatgtgcagcagtatcgtgc	819
QY	2329	aattcaactgtgaaagagtgtacacatacagcgtgtgtgtctgtacatgtgaaagtgtgatatgc	2388
Db	820	agttcaactgtgaaagagtgtacacatacagcgtgtgtgtgtgtgcctgtgaaagtgtgcctgc	879
QY	2389	aagctactcttgataaccggtacgttccaagctgtgtcgaccctagcagcgacatctccaaca	2448
Db	880	agcgcatcctctgacaacggcagcactccaagctgtgtcggcccgccagcgacatctccaaca	939
QY	2449	ttcagcaagcattatgggacacacacagaacacagatagcgtgtagtgttacaatagattgcgaca	2508
Db	940	ttcagcgagcgcatattgggacacacacagaacacagatagcgtgtagtgttacaatagcgtgcaca	999
QY	2509	accttaactcatalctccacacqttgtctcttgaagatcaacgcaagaatttaccacatgcaccc	2568

Db 1000 accctgagctacatgccactgtgtctttgagatcatatgcaaatgtaccactgacc 1059
Qy 2569 ccttcgcatataccagcagcagatcaagggtcttgcacccagtgatccagatggaacc 2628
Db 1060 ccttcgcatataccagcagcagatcaagggtcttgcacccagtgatccagatggaatc 1119
Qy 2629 attccagaatgtatcttggagatgttcatctgtagtactaacagctcttgaca 2688
Db 1120 attccagaatgtatcttggagatgttcatctgtagtactaacagctcttgaca 1179
Qy 2689 gggccacaacacctggtggcgtagctaaagcaatctgaagctt 2731
Db 1180 gggccacaacacctggtggcgtagctaaagcaatctgaagctt 1222

RESULT 9

AAN40295
ID AAN40295 standard; mRNA; 1175 BP.

AC AAN40295;

DT 04-FEB-1992 (first entry)

DE Sequence encoding a polypeptide displaying milk clotting activity.

EW Cheese-making: recombinant protein; rennet substitute; milk clot; ss.

OS Bos taurus.

FX Key Location/Qualifiers

FT sig_peptide 21..69

FT mat_peptide 70..1166

FT tag- b

PN EP123928-A.

XX 07-NOV-1984.

PF 30-MAR-1984; 84EP-0103551.

PR 31-MAR-1983; 83US-0480860.

PA (CODON-) CODON GENETIC ENG.

PI Cashion LM, McCaman MT, Rice CW, Sias SR;

DR WPI: 1984-277277/45.

XX P-PSDB: AAP40559.

PT Recombinant DNA coding for milk clotting polypeptide - which is expressed in transformed bacteria

PS Claim 6; Fig 2; 39pp; English.

CC Also claimed is E.coli JM83/pLC7 (ATCC 39325) which is transformed with pLC7 contg. the prorennin derived sequence fused in phase with B-galactosidase. The pLC7 prorennin expression plasmid includes

CC sequences which code for both the pseudorennin and mature rennin cleavage sites between AAs 28-29 and AAs 42-43, respectively.

XX Sequence 1175 BP; 263 A; 355 C; 319 G; 238 U; 0 other:

Query Match 22.7%; Score 897.2; DB 5; Length 1175;
Best Local Similarity 68.2%; Pred. No. 3.8e-118;
Matches 763; Conservative 217; Mismatches 138; Indels 0; Gaps 0;

Qy 1609 tctgtcgttactacagctgtgtagtaccgcatctcttctacaaagtgatctc 1668
Db 49 ucuucgucucuccagagcgucugagaucacccaggaucuccuacaaaggaagucuc 108

Qy 1669 tccgttaaggcgtgtgaaggaacatgtactcttagaagactcttgcagaaacaaacatgt 1728
Db 109 ugaggaaggcgucugaaggagcuaugggcuucugagagcauuccugcagaacacagauag 168
Qy 1729 gcatcagcagcaagatctccgcgttcgtgtaagtctgtatgctgtcgttaccataactac 1788
Db 169 gcaucagcagcaggaucucgc 228
Qy 1789 ttgatgtataatattcttggaagatcttaccctcgaaaccccgctcaagatgttaccgttc 1848
Db 229 uggaugagucacuuuuggaagaaucacucggagcccgcccgcccgaggaugacccugc 288
Qy 1849 tcttgatactgttccctgactcttgggttcctcccttactactactacagagcaagct 1908
Db 289 uguuugaacacucgucucucucgacuuucgguacccuacuaucugcaagagcaauccu 348
Qy 1909 gcaagaacacacaaagatctgcattcagacagacagtagtagttagcaccacgaagcaaac 1968
Db 349 gcaaaacacacagcgcuucgaccccgagaaagucgucacuuuccagaaacugcgcaagc 408
Qy 1969 ccttgctatacactcaggtacaggtacgtacgaagaatcttaggtatgataccgtca 2028
Db 409 ccugucuaucacuaucgagacagcagcaugcaugcagggcauccuagacuaugacacac 468
Qy 2029 cgtgtccacacattgtgtgacatccacagacagtagtagttagcaccacgaagcaagct 2088
Db 469 cuugucacacaaauugugacauccagagagacagaaugccuagacacaccccgaggg 528
Qy 2089 atgtcttaccctatgagaaattcgatgtacatccttgtagttagcaccatcgctcgt 2148
Db 529 acgucucacacuaugcggaauucgacgcggaucougugggauugcuaacccucugcgc 588
Qy 2149 cagagtagctcgtatcctgtgttgtaaacatgatgatgacccgaacactagctcaagat 2208
Db 589 cagagtagcagaaacccgucuuugacacaaugacacagacagcucugugcgcaagacc 648
Qy 2209 tgttcgcgtttacatgtgacaggaatgtgcagagagacagtagctcaagcttgagatgt 2268
Db 649 uguucucgcuuuaacuaugacaggaagacaggaagacagcuaucagcucugggcgcaug 708
Qy 2269 atccatcctactacacagatctctcactggttcagctcaactgtgcagcagtagctgc 2328
Db 709 acccgucuaucacacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 768
Qy 2329 aattcactgtgacagctgtgcacacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2388
Db 769 aguuacacugugacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 828
Qy 2389 aagctactctgtgataccggtacgtccaaagctgtgcagccttagcagcagcagcagc 2448
Db 829 aggcacauccugacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 888
Qy 2449 ttacgaagaactatgtgagcagacacagaacagtagcgtgtgtgtgtgtgtgtgtgt 2508
Db 889 uccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 948
Qy 2509 accctagcactactgtcctaaagtgtcttggagatacagcagcaagatgtccacagcacc 2568
Db 949 accuagacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1008
Qy 2569 ccttcgcatataccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2628
Db 1009 ccuucgcuuaucacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1068
Qy 2629 attccagaatgtatcttggagatgttcatctgtagtactaacagctcttgaca 2688
Db 1069 auuccagaatgtatcttggagatgttcatctgtagtactaacagctcttgaca 1128
Qy 2689 gggccacaacacctggtggcgtagctaaagcaatctga 2726
Db 1129 gggccacaacacctggtggcgtagctaaagcaatctga 1166

```
RESULT 10
AAN40180
AAN40180 standard; cDNA; 1460 BP.
XX
AC AAN40180;
XX
DT 25-JAN-1992 (first entry)
XX
DE Sequence of recombinant CGF4 carrying the rennin coding sequence.
XX
KM Yeast expression vector; GAL1 promoter; Saccharomyces cerevisiae;
XX
KW ss.
XX
OS Bos taurus.
XX
FT Key Location/Qualifiers
FT CDS 205..1350
FT /*tag= a
XX
PN GB2137208-A.
XX
PD 03-OCT-1984.
XX
PF 28-FEB-1984; 84GB-0405129.
XX
PR 28-FEB-1983; 83US-0470911.
XX
PA (COLB ) COLLABORATIVE RES INC.
XX
PI Botstein D, Davis RW, Fink GR, Taunton-Rigby A, Knowlton RG;
PI Mao JI, Moir DT, Golf CG;
PI WPI: 1984-245517/40.
DR P-PSDB; AAP40218.
XX
PT DNA segment confg. GAL1 promoter linked to gene - useful for
PT direction of expression of the gene in yeast cell
XX
PS Example; Table 4, Page 21-23; 35pp; English.
XX
CC The inventors claim a DNA segment confg. GAL1 promoter linked to
CC gene - useful for direction of expression of the gene in yeast cell.
CC The recombinant material carrying a GAL1 promoter of the yeast
CC galactokinase gene may be used in expressing a desired protein, esp.
CC bovine growth hormone, interferon, prorennin or preprorennin. In the
CC yeast cell. Strains of Saccharomyces cerevisiae producing the
CC polypeptides are produced. Yeast strains deposited as ATCC 20643,
CC 20661, 20662 and 20663, strain designations CGY 196, 457, 461 and
CC 528, resp. are new.
XX
SQ Sequence 1460 BP; 328 A; 440 C; 399 G; 293 T; 0 other;
XX
XX
Query Match 22.7%; Score 897.2; DB 5; Length 1460;
Best Local Similarity 87.7%; Pred. No. 3.8e-118;
Matches 980; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
OY 1609 tctgtgctgttactcagctgtgagatcacccgcatctctctctcaagaagtaagcttc 1668
DB 233 tcttgctctctcccgagcgctgagatcaccaagatccctctgttcaagaagttccc 292
OY 1669 tccgtaaagcgctgaagaacatgagctctctagaagaattcttgcaagaacaagtaag 1728
DB 293 tgaggaagcgctgaagaagcatgagctcttggaagattctctgcagaacaagcagtaag 352
OY 1729 gatacagcagcaagatctccgcttcgtggaagtgtgctgagcgctgcaacttacaactacc 1788
DB 353 gatacagcagcaagatctccgcttcgtggaagtgtgctgagcgctgcaacttacaactacc 412
OY 1789 ttgatagtcaataactcttggaagaatctactcgtgaaccccgccctcaagaagttcaccttc 1848
DB 413 tggatagtcaagtaactcttggaagaatctactcgtgaaccccgcccgagagttcaccttcg 472
```

```
OY 1849 tcttgtaactggttctcttgactcttgggttccctctatctatctgcaagaagtaagcct 1908
DB 473 tgttgacaactgctctctgactcttggttaccctctatctctgcaagaagtaagcct 532
OY 1909 gcaagaaccaccaagaatttcgatccggaagaagtgtccactccactccaacttaagcaaac 1968
DB 533 gcaaaaaccaccagctctgcagcccggaagaagtgtccactccactccaacttaagcaaac 592
OY 1969 cctgtctatacactacgtagcaagtagcaagtagcaagaatcttgactatgataccgtca 2028
DB 593 cccgtctatccactacgtagcaagtagcaagtagcaagaatcttgactatgataccgtca 652
OY 2029 ctgtctccaagaattgtagacttcaacagagatgagacttagacccaagaacagatg 2088
DB 653 ctgtctccaagaattgtagacttcaacagagatgagacttagacccaagaacagatg 712
OY 2089 atgtctcaactatgcaagaattcgatgcactctgtgatgtagtataccactcgtcgct 2148
DB 713 acgtcttcaactatgcaagaattcgatgcactctgtgatgtagtataccactcgtcgct 772
OY 2149 cagaagtaactgatacctgtgttgacaacatgatagaacgcgaacactagtaactaagaact 2208
DB 773 cagaagtaactgatacctgtgttgacaacatgatagaacgcgaacactagtaactaagaact 832
OY 2209 tgttctcggtttacatgagcaaggaaatggccaggaagacatgcaacgcttggaactatg 2268
DB 833 tgttctcggtttacatgagcaaggaaatggccaggaagacatgcaacgcttggaactatg 892
OY 2269 atccatctactacacagaagatctctcactgtgtctcagtaactgtgcaagaactatg 2328
DB 893 accgctctactacacagaagatctctcactgtgtctcagtaactgtgcaagaactatg 952
OY 2329 aattcaactgtggaacagtgtcaccaatcagaacggtgtgtgtgtgacatgtaagtgtagtc 2388
DB 953 agtcaactgtggaacagtgtcaccaatcagaacggtgtgtgtgtgacatgtaagtgtagtc 1012
OY 2389 aagctatcttgatataccggtacagctgtcgaagctgtcgaactcagcgacatcttcaaca 2448
DB 1013 aagcctatcttgatataccggtacagctgtcgaagctgtcgaactcagcgacatcttcaaca 1072
OY 2449 ttcaagaagctatgtggaagcacacagaacacagtaacgtgagtttgaacatagattgcaga 2508
DB 1073 tccagcagcgcatgtggaagcacacagaacacagtaacgtgagtttgaacatagattgcaga 1132
OY 2509 accttaagtaactgcttaagaagtgtcttgagataaagcgaagtgtaaccactgagacc 2568
DB 1133 accttaagtaactgcttaagaagtgtcttgagataaagcgaagtgtaaccactgagacc 1192
OY 2569 ccttcgactataccagcagatcagaaggttctgcaaccagtgagatccagagtgagaacc 2628
DB 1193 ccttcgactataccagcagatcagaaggttctgcaaccagtgagatccagagtgagaacc 1252
OY 2629 attccagaatgagatcttggaagatggtgttcaattcgtgagtaactaagcgtcttgaga 2688
DB 1253 attccagaatgagatcttggaagatggtgttcaattcgtgagtaactaagcgtcttgaga 1312
OY 2689 gggccaacaacactcgttgggtgagctagctaaagcaatcga 2726
DB 1313 gggccaacaacactcgttgggtgagctagctaaagcaatcga 1350
RESULT 11
AAQ14051
AAQ14051 standard; DNA; 1210 BP.
XX
AC AAQ14051;
XX
DT 06-JAN-1992 (first entry)
XX
KM Rennin gene.
DE Prorennin; alpha-S1-casein gene; insulin-like growth factor I; IGF-I;
KW mammary gland; ss.
```

XX OS Synthetic.
 XX PN EP451823-A.
 XX PD 16-OCT-1991.
 XX PF 10-APR-1991; 91EP-0105702.
 XX PR 19-APR-1990; 90DE-4012526.
 XX PR 11-APR-1990; 90DE-4011751.
 XX PA (CONSORT ELEKTROCHEM IND.
 XX PI Hartl P, Brem G;
 XX DR WPI; 1991-304858/42.
 XX PT Recombinant DNA constructs for expressing protein in milk -
 XX PT contg. specific mammary gland transcription control region and
 XX PT signal sequence, providing high yield and easy prod. recovery
 XX PS Disclosure; Page 21; 41pp; German.
 XX CC The remnin gene was isolated using the probe represented in AA014775.
 XX CC It was used as heterologous peptide/protein together with parts of the
 XX CC alpha-SI-casein gene in the prodn. of DNA constructs. The heterologous
 XX CC peptide or protein may also be human insulin-like growth factor I.
 XX CC The constructs provide high yields of the protein with simple recovery
 XX CC from the milk. Activation of the gene occurs only in the mammary gland.
 XX CC See also AA014050, AA014774-77.
 XX SC Sequence 1210 BP; 278 A; 356 C; 324 G; 252 T; 0 other;
 Query Match 22.7%; Score 896.4; DB 12; Length 1210;
 Best Local Similarity 85.7%; Pred. No. 4.9e-118;
 Matches 996; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 1609 tccttggctgactcaagcgtgctgagatcccccattccctctctacaagaagtaagcttc 1668
 DB 43 ttgtggtctgttgcctcttcgacacgagatccacagaatccctctgtacaagaagcttc 102
 QY 1669 tccgttaaggcgtctgaagaagaatgactcttagaagaactcttgacagaagaacagatag 1728
 DB 103 tgaaggaaggcgtctgaagaagcagatggcttctggaagactcttcgcaaaaagaagatag 162
 QY 1729 gcatcagcagcaagctactccggctcgggtgaagtctgaagctgacacttaaccaataacc 1788
 DB 163 gcatcagcagcaagctactccggctcgggaagtgagcgaagcgtgcccctgacccaataacc 222
 QY 1789 ttatataagtaactcttgggaagaatctactcctcgaaacccgcctcaagaagctaccgttc 1848
 DB 223 tgaataagctgactcttgggaagaatctactcctcggaacccgcctcaagaagctaccgttc 282
 QY 1849 tcttctgaactggttccctcgactcttggttccctctatctacttgaagaagaactcct 1908
 DB 283 tgtttgcaactggtcctctcgactcttggttaccctctatctacttgaagaagaactcct 342
 QY 1909 gcaagaacccaagaagaatcgatccgaagaagtcgtccaccttcagaagaacttaggaac 1968
 DB 343 gcaaaaacccaagcgtctcgacccgagaagaagtcgtccaccttcagaagaacttaggaac 402
 QY 1969 cctgtctctacactagcgtgacagtagcagtagcagaagaacttaggtatgatacgtca 2028
 DB 403 cctgtctctacactagcgtgacagtagcagtagcagtagcagtagcagtagcagtagcag 462
 QY 2029 ctgtctctcaacatgttgacaattacaacagagtaggacttagcaccagaagaacaggtg 2088
 DB 463 ctgtctctcaacatgttgacaattacaacagagtaggacttagcaccagaagaacaggtg 522
 QY 2089 atgtctctacactatgcagaatctcgatgaccttggatatgacataccatcgctcgct 2148

DB 523 acgtcttcaacctatgcccgaatttgacggaatcctcgggagtgacctaccctcgctcgct 582
 QY 2149 cagagtaactgataacctgtgtttgacaacatgatagtaaacacactagtagtaagaact 2208
 DB 583 cagaagtaactgataacctgtgtttgacaacatgatagtaaacacactagtagtaagaact 642
 QY 2209 tgttctcggtttatattgacagagaatgcccagagagaagatgctcaagcttgaagctatg 2268
 DB 643 tgtctcgtttatattgacagagaatgcccagagagaagatgctcaagcttgaagctatg 702
 QY 2269 atccatcctactacaagaagatctcttaactgtgttccaagtaactgtgacagtaactggc 2328
 DB 703 accgctcctactacaagaagatctcttaactgtgttccaagtaactgtgacagtaactggc 762
 QY 2329 aattcacgtgtgacagatgctccatcagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2388
 DB 763 agttcaactgtgtgacagatgctccatcagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 822
 QY 2389 aagctatcttggataccggtacgttccaagctgtgtcgaagcttagcagagacattctcaaca 2448
 DB 823 aggcatactctgacacagcggcaccctccaagctgtgtcggccagcagagatctcctcaaca 882
 QY 2449 ttcaagcaagctatgtgagcacaacagaacacagatcagtgagtttgacatagatgagaca 2508
 DB 883 tccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 942
 QY 2509 acctagctataccctcacaagttgtctttgagatcaacagagaagatgtaaccactgagccc 2568
 DB 943 accggaagctataccctcacaagttgtctttgagatcaacagagaagatgtaaccactgagccc 1002
 QY 2569 cctcgcctataccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2628
 DB 1003 cctcgcctataccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1062
 QY 2629 attcccaagaatgagatcttggagagatgtgtcattcgttgactacacagcgtctttgaca 2688
 DB 1063 attcccaagaatgagatcttggagagatgtgtcattcgttgactacacagcgtctttgaca 1122
 QY 2689 gggcacaacactcgttggggtagctaaagaactgaagcttaagttgaagttgaactaa 2748
 DB 1123 gggcacaacactcgttggggtagctaaagaactgaagcttaagttgaagttgaactaa 2782
 QY 2749 atgcattgaggtgaagagctc 2770
 DB 1183 gcgtcgcacaagcttgagagagctc 1204

RESULT 12
 AAZ06463
 ID AAZ06463 standard; DNA: 2733 BP.
 XX
 AC AAZ06463;
 XX
 DT 29-NOV-1999 (first entry)
 XX
 DE 2.7 Kbp HindIII fragment of pBSOTPRNT (oleosin-chymosin fusion gene).
 XX
 KW oll-body; lipid body; oleosome; spherosome; separation;
 KW fusion protein; heterologous polypeptide; commercial production;
 KW plasmid; ds.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT 850..2729
 FT CDS
 FT /*tag= a
 FT /product= "oleosin-spacer-Met-prochymosin"
 FT /note= "CDS contains an intron"
 FT exon 850..1202
 FT /number= 1
 FT intron 1203..1442
 FT /number= 1
 FT exon 1443..2733

FT XX /*tag= b
PN EPI21775-A.
XX 17-OCT-1984.
PD
XX 07-MAR-1984; 84EP-0102451.
XX
PR 09-MAR-1983; 83JP-0038439.
XX
PA (BEPP/) BEPPU T.
XX
PI Beppu T, Uozumi T, Nishimori K, Shimizu N, Kawaguchi Y;
PI Hidaka M;
DR WPI: 1984-258001/42.
DR P-PSDB: AAP40078.
XX
PT Expression plasmid comprising prochymosin gene and vector -
PT useful for transforming Escherichia coli for prochymosin prodn.
XX
PS Disclosure: Fig 1; 59pp; English.
XX
CC The inventors claim the prochymosin gene comprising a nucleotide
CC sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or
CC (b) the 5th codon (CGG) to the 365th codon (ATC); and recombinant
CC plasmids harboured by Escherichia coli strains deposited as FERM BP-
CC 262, -263 and -264. Any portion of the nucleotide sequence as
CC described in AAN40055 can be used. Also claimed is a vector derived
CC from plasmid pBR322. Typically plasmid pBR301 is obtd. from pOCT 2.
CC The transcriptional direction of pOCT 3 is opposite to that of
CC pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter
CC clockwise in pOCT 3.
XX
SQ Sequence 1278 BP; 309 A; 384 C; 326 G; 259 T; 0 other;

Query Match 22.6%; Score 895.6; DB 5; Length 1278;
Best Local Similarity 87.6%; Pred. No. 6.4e-118;
Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 1609 tcgtgtgttattcaccagctctgagatccgcgaattcctctctcaagaagtgatctc 1668
DB 13 tcttcgctctcccgagcgctgagatccagagatccctctgtacaagcgaaagtc 72

QY 1669 tccgtaagcgctgaaggaatgactcttgaagaactcttctcagaagaacagatg 1728
DB 73 tgaagaaagcgctgaaggaagcagtggtctctcggaagactcttcgaagaacagatg 132

QY 1729 gcatcagcagaagctactccggtctcggtgaagtgtagctgacacttaccactacc 1788
DB 133 gcatcagcagaagctactccggtctcggtgaggtgagcgagtgccctgacactacc 192

QY 1789 ttgatagtcaatattctggggaagatctaccctcgaaacccgcctccaagatcacgctc 1848
DB 193 tggatagctagctactctgggaagatctaccctcgaaacccgcctccaagatcacgctc 252

QY 1849 tcttgatactggtctctactctggtctcggtctctactctacttactgcaagcgatgct 1908
DB 253 tgtttgacactggtctctctctgactctctggtaccccttacttactgcaagcgatgct 312

QY 1909 gcaagaacccaagaatctgactcgaagaagtcgctcacccttcagaaactttaggcaaac 1968
DB 313 gcaaaacacacagcgctctgacccggaagaagtcgtctcacttcagaaactttaggcaaac 372

QY 1969 cctgtctatatactatagtgatcagtgatcagaagaaatcttaggtctatgataccgtca 2028
DB 373 cccgtctatatactatagtgatcagaagaaatcttaggtctatgataccgtca 432

QY 2029 ctgtctccaacatgttgatactcaacagacagctaggaacttagcaacccaagcagtg 2088
DB 433 ctgtctccaacatgttgatactcaacagacagctaggaacttagcaacccaagcagtg 492

QY 2089 atgtctcactatgcaagaattcgatggacatctgtgatatgcataccatccatcgctcgt 2148
DB 493 agctctcactatgcaagaattcgatggacatctgtgatatgcataccatccatcgctcgt 552

QY 2149 cagagctactgatactctgttttgacaacatgatacagcaacacttagtactgaagact 2208
DB 553 cagagctactgatactctgttttgacaacatgatacagcaacacttagtactgaagact 612

QY 2209 tttcttggttttatactgagcaggaatggcgggaagagatgtctcaagcttgagttatg 2268
DB 613 tttcttggttttatactgagcaggaatggcgggaagagatgtctcaagcttgagttatg 672

QY 2269 atccatctactacacagaagatctctcactgtggtctcagctactgtgagcagatctgac 2328
DB 673 accgctctactacacagaagatctctcactgtggtctcagctactgtgagcagatctgac 732

QY 2329 aattcactgtgacagatgtcaacatcagcagtggtgtgtgtgtgcatgtaaggtgagtg 2388
DB 733 agttcactgtgacagatgtcaacatcagcagtggtgtgtgtgtgcatgtaaggtgagtg 792

QY 2389 aagctatcttgatccggtgacagctggtcgcgcagctagcagcagatcttcaaca 2448
DB 793 aggcatacttgacagcgggacactccaaagctggtcgcgcagcagcagatcttcaaca 852

QY 2449 ttcaagcagctatgtgagccacaacagaacagctagcgtgagttgacataagatcgaca 2508
DB 853 ttcaagcagctatgtgagccacaacagaacagctagcgtgagttgacataagatcgaca 912

QY 2509 accttagctatactgctacagtgctcttgagatcaacggcaagaatgtaccactgaccc 2568
DB 913 accttagctatactgctacagtgctcttgagatcaacggcaagaatgtaccactgaccc 972

QY 2569 ccttcgcctatataccagcaggaatgaaggtctctgacacagcggatctcagaagttgagaac 2628
DB 973 ccttcgcctatataccagcaggaatgaaggtctctgacacagcggatctcagaagttgagaac 1032

QY 2629 attccagaatgtgactctggagagatgcttcacatctcgtgaagctactacagcgctcttgaca 2688
DB 1033 attccagaatgtgactctggagagatgcttcacatctcgtgaagctactacagcgctcttgaca 1092

QY 2689 gggccaacaacctggttggtgctagctaaagcaatctga 2726
DB 1093 gggccaacaacctggttggtgctagctaaagcaatctga 1130

RESULT 14
AAN20043
ID AAN20043 standard; DNA; 1460 BP.
XX
AC AAN20043;
XX
DT 16-DEC-1992 (first entry)
XX
DE Pre-prorennin-A gene DNA sequence.
XX
KW Pre-pro-rennin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;
KW protease; milk-clotting enzyme; ss.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT CDS 205..1350
FT /*tag= a
FT /label= a
FT /note= "pre-prorennin-A gene"

GB2091271-A.
28-JUL-1982.
15-JAN-1982; 82GB-0001120.
01-DEC-1981; 81US-0325481.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 15:44:34 ; Search time 5848.1 Seconds
(without alignments)
9132.445 Million cell updates/sec

Title: US-09-643-755B-3
Perfect score: 3957
Sequence: 1 ccgcaggaatcatgtact.....accctaccactaaggtacc 3957

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estdb:*
2: em_esthm:*
3: em_estln:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estc:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vtc:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	387.2	9.8	472	10	BG938320 1AB015E12
2	358	9.0	631	9	A1892188 mjb4c05.y
3	354.8	9.0	430	10	BG937697 1AB005E01
4	337.2	8.5	727	9	A1326975 mjb1e03.x
5	333.8	8.2	399	10	BG937723 1AB005D06
6	301.4	7.6	1347	11	AK004109 Mus muscu
7	300.6	7.6	383	10	BG938086 1AB011A08
8	262	6.6	479	9	AA028632 m14c08.r
9	249.6	6.3	505	9	A1479358 tm27e07.x
10	246	6.2	546	9	A1324867 m18b01.x
11	242.4	6.1	517	9	A1385490 mjb1e03.y
12	236.2	6.0	1385	11	AK008959 Mus muscu
13	236	6.0	619	9	AM868692 MRL-SN06
14	227.4	5.7	1388	11	AK008886 Mus muscu
15	226.8	5.7	414	9	A1322423 m18b01.y
16	217.8	5.5	636	9	AM868716 MRL-SN06
17	206.4	5.2	814	10	B1761345 603043918

18	205.8	5.2	583	10	BE841742	BE841742 MRL-SN006
19	204.4	5.2	910	10	BF299798	BF299798 602030421
20	202.8	5.1	622	10	BF373831	BF373831 MRL-SN000
21	201	5.1	716	9	AM012992	AM012992 SSERN01SK
22	199.4	5.0	603	10	BE841740	BE841740 MRL-SN006
23	199	5.0	702	9	AM867433	AM867433 MRL-SN003
24	198.6	5.0	551	10	BE841714	BE841714 MRL-SN006
25	196.4	5.0	547	10	BM069200	BM069200 id71a02.y
26	196.4	5.0	548	10	BM053683	BM053683 id69b07.y
27	190.6	4.8	391	9	AA969042	AA969042 op43d07.s
28	190	4.8	433	10	W34201	W34201 ma99h09.r1
29	190	4.8	713	10	BF121716	BF121716 601759230
30	187.4	4.7	422	10	W10274	W10274 ma37g09.r1
31	185	4.7	666	10	BE841761	BE841761 MRL-SN006
32	182.6	4.6	594	9	AM863768	AM863768 MRL-SN001
33	182.2	4.6	578	10	BE840800	BE840800 MRL-SN000
34	182	4.6	564	9	AM868711	AM868711 MRL-SN006
35	181	4.6	736	10	B1821121	B1821121 603035596
36	181	4.6	847	10	B1490133	B1490133 603031966
37	180.8	4.6	616	10	BE841758	BE841758 MRL-SN006
38	180.4	4.6	547	10	BE841657	BE841657 MRL-SN006
39	180.4	4.6	576	9	AM868673	AM868673 MRL-SN006
40	179.6	4.5	960	10	B1517784	B1517784 603042031
41	179.4	4.5	751	9	A1076336	A1076336 oz04a06.x
42	178.8	4.5	815	10	B1818940	B1818940 603037434
43	176.4	4.5	666	9	AM959843	AM959843 EST372014
44	176.2	4.5	847	10	B1762154	B1762154 603049340
45	176	4.4	572	10	BF373812	BF373812 MRL-SN000

ALIGNMENTS

RESULT 1
BG938320 472 bp mRNA linear EST 11-JUN-2001
LOCUS 1AB015E12 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
DEFINITION sequence.
ACCESSION BG938320.1 GI:14337692
VERSION BG938320.1
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae; Bovidae; Bos.
REFERENCE 1 (bases 1 to 472)
AUTHORS Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
TITLE CDNA's from bovine abomasum tissue
JOURNAL unpublished (2001)
COMMENT Contact: Dr. Stephen Moore
. Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin) mrna) in main database at high score of 928.0 and E-value of 0.0
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 472
POLYA-No.
FEATURES
source location/Qualifiers
1..472
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bovine Abomasum cDNA Library"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"
/cell_type="Epithelial"

DB	541	AAATTGGGACGTTACAGTGGACAGGATGCATCATCATGTGAAGTGGTGGTTCACAGG	600
QY	2381	tggtatgtcaagctatcttggataccggtac	2410
Db	601	TGGCTGCGCCCTGGCGTTCTGGACACAGGAC	630
RESULT	3		
LOCUS	BG937697	430 bp	mRNA
DEFINITION	LA0005E01	Bovine Abomasum cDNA Library	Bos taurus cDNA 5', mRNA
ACCESSION	BG937697		
VERSION	BG937697.1	GI:14337069	
KEYWORDS	EST.		
SOURCE	COW.		
ORGANISM	Bos taurus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
	Bovidae; Bovinae; Bos.		
	1 (bases 1 to 430)		
REFERENCE	Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.		
AUTHORS	CDNA's from bovine abomasum tissue		
TITLE	Unpublished (2001)		
JOURNAL	Contact: Dr. Stephen Moore		
COMMENT	. Beef Genomics Laboratory		
	Dept of AFNS, University of Alberta		
	410 Agri/For. Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada		
	Tel: 780 492 0169		
	Fax: 780 492 4265		
	Email: smoores@afns.ualberta.ca		
	The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin)		
	mRNA) in main database at high score of 844.0 and E-value of 0.0		
	PCR primers		
	FORWARD: M13 Forward		
	BACKWARD: M13 Reverse		
	Seq primer: T3 primer		
	High quality sequence stop: 430		
	POLYA-NO.		
FEATURES			
source	Location/Qualifiers		
	1..430		
	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
	/clone_lib="Bovine Abomasum cDNA Library"		
	/sex="Two males and one female mixed"		
	/tissue_type="Gastrointestinal tissue (GIT)"		
	/cell_type="Epithelial"		
	/dev_stage="Young adult"		
	/lab_host="XLI-BlueMRF"-strain"		
	/note="Organ: Abomasum; Vector: Uni-22APXR; Site_1: Ecor		
	I; Site_2: Xho I"		
BASE COUNT	97 a 130 c 119 g 84 t		
ORIGIN			
Query Match	9.0%; Score 354.8; DB 10; Length 430;		
Best Local Similarity	89.1%; Pred. No. 1.9e-36;		
Matches 383; Conservative	0; Mismatches 47; Indels 0; Gaps 0;		
QY	2158	cgataccgtgttggacaacatgatgatccgacacactgtagcttaagaacttgtctcgg	2217
Db	1	CGATRACCGGTGTTGACACATGATGAACAGGACACCTGGTGGCCCAAGACCTGTTCTCGG	60
QY	2218	tttataatgacaggaatggccagagagacatgctcaagcttgagactttaacatcat	2277
Db	61	TTTACATGGACACGGAATGGCCAGGAGACATGCTACGCTGGGGCCATTCGACCCGTCCT	120
QY	2278	actacacagatctcttctacttgggttccagttacgttcagtcagagacttgcgaatcacty	2337
Db	121	ACTACACAGAGGTCCTGCACACTGGGGCCCGTACAGTGGAGCAGTACTGGCAGTTACAGC	180
QY	2338	tggacagtgtcaaccatcagcgggtgtgtgttgttgcattgtgaagtgatgtcaagctatct	2397

Db	181	TGGACAGGTCAACCATCAGCGGCTGTGTGGTTCGTGGCCCTGTGAGGGTGCGTGTACAGGCCATCC	240
OY	2398	tggataccggtacgtccaaagctgtgcggacctagcagcgacattctcaacattcagaag	2457
Db	241	TGGACACGCGGCACCTCCAAAGCTGTGTCGGGGCCAGCAGCGACATCCTCAACATCCAGAGG	300
OY	2458	ctatggaggccaacaagaacagttacggttgagttagtgaatagatattggacaacctagct	2517
Db	301	CCATTGGAGCCACAACAGAACAGTAGATGAGTTTGACATGCACACTGCCACAACCTCAGACT	360
OY	2518	aattgccctcacggttcctttgatattcaacgcaagatbtaccacttgcccccttcgcct	2577
Db	361	ACATGCCCACACTGTGCTCTTTGAGATCAATGGCAAATGTACCACCTAGCACCCCTCGCCT	420
OY	2578	ataccagcca	2587
Db	421	ATACAGCCA	430
RESULT	4	A1326975	727 bp mRNA linear EST 23-DEC-1998
LOCUS	A1326975/c	m81e03.x1 Soares mouse p3NMF19.5 Mus musculus cDNA clone	
DEFINITION		IMAGE:483524 3' similar to TR:Q28075 Q28075 CHROMOSIN C.; ; mRNA sequence.	
ACCESSION	A1326975		
VERSION	A1326975		
KEYWORDS	A1326975.1 GI:4061404		
SOURCE	EST.		
ORGANISM	house mouse.		
	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 727)		
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The WashU-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:293268 This clone was previously sequenced on the 5' end only, this new data is from the 3' end Possible reversed clone: similarity on wrong strand High quality sequence stop: 446. Location/Qualifiers 1..727 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:482524" /clone_id="Soares mouse p3NMF19.5" /dev_stage="19.5 dpc total fetus" /lab_host="DH10B (ampicillin resistant)" /note="Vector: pUT7R3D (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTATCCAAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pUT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."		
FEATURES	source		

Db	Accession	Gene	Species	Length	EST	EST Date
Db	700	AAAGGACAGCATGATGCTAGTGGAGTGGATGATCCCTCTACTACCATGAGAGCTTAC	Bos taurus	759		
Qy	2238	tggatccagtcactctgycagcagagtiactgtgcaattcaactctgtggacagtgatccatcagc	Bos taurus	2357		
Db	760	TGGGTACCAAGTGTCCAAACCCAGCTACAGGCAATTAAGTGTGATGATGATTCATGTAAT	Bos taurus	819		
Qy	2358	ggtgtgtgtgtgtgatgtgaagtgatgtcaagctatcttgatlaaccgtgacgtcgaag	Bos taurus	2417		
Db	820	GGGAGAGCATTTGCTGCTGTGATGTGTGGCTGCGCCAAAGTATTATGACACAGGAGACCTCCTTG	Bos taurus	879		
Qy	2418	ctgtgtcgagacctagcagcgcacattctcaacatlaagcaagctatltgagccacacagaac	Bos taurus	2477		
Db	880	CTGACCGGCCCCGCAAGGCTCATGCTTAACATTCACAGATCTAATGTGGTCCAAAGGCTTCT	Bos taurus	939		
Qy	2478	cagtaacggtgtgatttgatatgtatgtgcacaaccttagctacaggtccacagtgctctt	Bos taurus	2537		
Db	940	GGTACGGCGAGTACTTCTCAAGTGTACACACATCAACACCCCTGCTGATATTGTCTTC	Bos taurus	999		
Qy	2538	gagatcaacagcgaagatgtccacacttgacccccctccgctcattacacagccagatcaagg	Bos taurus	2597		
Db	1000	ACCATCGGCACATGTTTACTACCTACCAAGTGCACAGCAATGCTCATATCCGAAGATCGTCA	Bos taurus	1059		
Qy	2558	tctgcaccagtgatgtatccagaagtgtgaac-----cattccagaatg	Bos taurus	2642		
Db	1060	CACAAATTCACAGAGCAMACTTTGAGGAGGCGCATGTGATGCCATCAGACCTGAGTGTGG	Bos taurus	1119		
Qy	2643	atcttggagatgtgtttatctcgtgtgactacacagcgtctttgacagggccaacatcct	Bos taurus	2702		
Db	1120	GTGCTGGGGGATGTCTTCTCTGAGGCTGTATTTCACCCGTGTTTGATCGGGCAATATACAG	Bos taurus	1179		
Qy	2703	gttgggtactgactaaagcaatctgaaagcttaa	Bos taurus	2733		
Db	1180	ATTGGTCTGGCTCTCTGTCATGAGTGTGGA	Bos taurus	1210		

RESULT 7

RG938086 383 bp mRNA linear EST 11-JUN-2001

LOCUS L001A08 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA sequence.

ACCESSION RG938086

VERSION RG938086.1 GI:14337458

KEYWORDS EST.

SOURCE COW.

ORGANISM Bos taurus

EUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.

1 (bases 1 to 383)

Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,X. and Li,G.

cDNA's from bovine abomasum tissue unpublished (2001)

Contact: Dr. Stephen Moore

. Beef Genomics Laboratory

Dept of AFNS, University of Alberta

410 ARI/101, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada

Tel: 780 492 0169

Fax: 780 492 4265

Email: smoores@afns.ualberta.ca

The sequence best matches gb:BPUI9786 (Bos primigenius prothymosin mRNA, complete cds) in main database at high score of 735.0 and E-value of 0.0

PCR PRIMERS

FORWARD: M13 Forward

BACKWARD: M13 Reverse

Seq primer: T3 primer

High quality sequence stop: 383

E-VALUE=No

LOCATION/Qualifiers

1..383

/organism="Bos taurus"

/db_xref="taxon:9913"

Query Match	7.6%;	Score 300.6;	DB 10;	Length 383;
Best Local Similarity	88.3%;	Pred. No. 1.8e-29;		
Matches 338;	Conservative 0;	Mismatches 44;	Indels 1;	Gaps 1;
Db 1	AGTACTGATACCGCGTTTGACAAACATGATGAAACAGGACCTGGTGGCCCAAGACCTGT	60		
Qy 2152	agtactgcataccctgtgtttgacaacaatgatgaacccagacactagtagtcaagacttgt	2211		
Db 61	TCTCGGTTTACATGACGACGAAATGGCCAGAGAGCATCTCACCGTGGGGGCCATCGACCC	120		
Qy 2212	tctcggtttacatgtgaacaggaatgcccagagagacatgctcaagcttggagctatgtatc	2271		
Db 2272	catcctactaacacaggaatctcttcaactgtgttccaagtcacgtgtgagcaagtaactgcaat	2331		
Qy 2332	tcaactgtgacagtgatcaacatcaacagcggtgtgtgttctgtatgtgaagttgagtatcaag	2391		
Db 181	TCACGTGTGACAGTGTCAACATCACAGCGGTGTGTGTGTGTGGCTGTAGGGTGGCTGTGAGG	240		
Qy 2392	ctatcttgatataccggtacgcgttccaaagctgtgtgtgcagactagaacagatcttcaacatlc	2451		
Db 241	CCATCTCTGACACGCGGACCTCCAGCTGCGTGGGCGCCAGACGACACATCTCCAAACATCC	300		
Qy 2452	agcaagcattatggagacacacacagaaacacagctggtgagtttgatcat-agattggacaaac	2510		
Db 301	ACCAAGGCCATTTGGACGCCACACAGAAACCGATGAGGTGATGACATCTGACTGCGACACAC	360		
Qy 2511	cttagctacatgcccacacagttgt	2533		
Db 361	CTGAGCTACTATGCCACACTGTGCT	383		
RESULT 8				
LOCUS	AA028632	479 bp	mRNA	linear EST 11-SEP-1996
DEFINITION	ml14c08 r1 Soares mouse p3M9F19.5 Mus musculus cDNA clone IMAGE:463502.5, similar to SW:CHYM_BOVIN P00794 PROCHYMOSIN A AND B PRECURSORS ;, mRNA sequence.			
ACCESSION	AA028632			
VERSION	AA028632.1	GI:1494769		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus. 1 (bases 1 to 479)			
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.			
TITLE	The WashU-HHMI Mouse EST Project			
JOURNAL	Unpublished (1996)			
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu			

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

MI:277318

Possible reversed clone: similarity on wrong strand

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 444.

FEATURES

source

1. .479
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:463502"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I -oligo(dT) primer [5'
TGTTCACATCTGAGTGGAGCGGCGCATTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."

BASE COUNT

104 a 127 c 131 g 117 t

ORIGIN

Query Match

6.6%; Score 262; DB 9; Length 479;

Best Local Similarity 71.8%; Pred. No. 1.4e-24;

Matches 343; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 1963 gcaaacctgtctatactactaggtacaggtacgaaggaatttggtatgata 2022
|||||
Db 1 GCAAGCCCTGTTGTGTCAGTATGACTGTGTAAGTGAAGGCTTCTGGCTCAGACA 60
QY 2023 ccgtcactgtctccaacatgttgacattcaacagacagtagaacttagacccaagaac 2082
|||||
Db 61 CTGTACAGCTCTGATGATTTGATGCCATCATGACTGTGGGCTCAGTACAGAGAAC 120
QY 2083 caggtgagtcctcaacctatgacagaaattcgaatgacatccttgatagcataccatcgc 2142
|||||
Db 121 CGGGCGACATCTTCACCTACTCTCCATTTGATGATGGCATCTGGGCTGCTATCTCACTT 180
QY 2143 tcgcgtcagagtaactgatactacccgtgtgtttacacatgataagaccagacctagtagtc 2202
|||||
Db 181 TTGCTTCACATATCTCAGTACCAATATTGACACATGATGAACAGGCACTGGTGCC 240
QY 2203 aagactgtctcgtttatatacgaacagatgcccagagacatctcaagctttgag 2262
|||||
Db 241 AAGACGCTGCTCCGTTTACATGACGAGATAGACGAGGAGAGATCTCAGACTGGGGG 300
QY 2263 ctattgacatcactactacacagatctctcaactgtgttcaagtaactgtgcaagact 2322
|||||
Db 301 CCATGCTAGTCTACTTCTATAGGCTCAGTCACTGGTGGCTGTGACATGACAGGAT 360
QY 2323 acagggaattcactgttggaacagtgtaaccatcagcggtgtgtgttgcattgaaagt 2382
|||||
Db 361 ATTGGCAGTTCACAGTGAAGATGACATCAATGATGAGTGAAGTGGCTTGTCAGAGTG 420
QY 2383 gatgtcaagatccttgatcacccgttccaaagctgtgacacttagacagagact 2440
|||||
Db 421 GCGGCCCTGCCGTTTGAGACAGGCACTGCGCTGTGGAGGGGCTGTGATGAGACAT 478

RESULT 9
AI479358 505 bp mRNA linear EST 14-APR-1999
LOCUS AI479358/c
DEFINITION tm27e07.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2157828 3' similar to tr:028950 Q28950 PREPROCHYMOSIN
PRECURSOR ; , mRNA sequence.
ACCESSION AI479358

VERSION AI479358.1 GI:4372526

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 505)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index
Unpublished (1997)

JOURNAL

Contact: Robert Strausberg, Ph.D.

COMMENT

Email: cgaps-remail.nih.gov

This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 608 Std Error: 0.00

Seq primer: -400P from Glibco

High quality sequence stop: 367.

FEATURES

source

1. .505

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2157828"

/clone_lib="Soares_NFL_T_GBC_S1"

/lab_host="DH10B"

/note="Organ: pooled: Vector: pT73D-Pac (Pharmacia) with
a modified polylinker. Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBH19W, testis NT, and B-cell
NCI-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 257480-302087, 682632-687239,
726408-728711, and 729096-733399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

BASE COUNT 106 a 138 c 141 g 119 t 1 others

ORIGIN

Query Match

6.3%; Score 249.6; DB 9; Length 505;

Best Local Similarity 77.1%; Pred. No. 5.5e-23;

Matches 303; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 2234 actgtgacagtgccaccatcagcgtgtgtgtgtctcatgtgaagtgatgcaagct 2393
|||||
Db 505 ACCCTTNAAGTGCATATGACGCGGTGTGTGCTGTGACGGTGGCTGTCAAGGCC 446
QY 2394 atcttgataccggtatagctgaacgtgtgtgacctagcagcagcatctcaacattcag 2453
|||||
Db 445 ATCTTGACACCGGCACTCTCTGCTGTGTGGCTGTGGCAATCTTCAATCAG 386
QY 2454 caagctattggaacacacaggaaccagtaacgltgattgacatagattgcgaacact 2513
|||||
Db 385 CAGGCCATTGGAGCCACTGTGGGGCCAGTACATGATTGACATCGACTGGGGCGCTG 326
QY 2514 agctaatgacctacagttgtctttgagataaagggaagatgtacacactgacccctcc 2573
|||||
Db 325 AGCAGCATTTCCACAGCGCTGTTTGAATGCACACGCAAGAGTACCCCTGCCACCTCC 266
QY 2574 gccctatccagcaggaataagggttctgaacacagttgatttcagaaatcctc 2633
|||||
Db 265 GCTTATACCAAGCCAGGACCAAGGCTTTTGACCAAGTGTTCACAGGCTGACTATGATTCC 206
QY 2634 cagaatgtgactcttgagagatgtgttcacgttgtagtactacaagcgtctttgacagggcc 2693
|||||
Db 205 CAGCAGTGTGATCTCTGGGGAATGCTTTCATCTGGAGATTAACATGCTTTGACAGGACC 146
QY 2694 aacaactcgtttggtcagtaagaactctga 2726
|||||
Db 145 AATAACCGTGTGGGCTGCGAAGGCTGTCTGA 113

RESULT 10
 A1324867/c 546 bp mRNA linear EST 23-DEC-1998
 LOCUS A1324867
 DEFINITION m182801.x1 Soares mouse p3NMF19.5 mus musculus cDNA clone
 IMAGE:473065 3' similar to TR:Q28075 Q28075 CHYMOSIN C.; mRNA
 sequence.
 ACCESSION A1324867
 VERSION A1324867.1 GI:4059296
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 546)
 REFERENCE
 AUTHORS Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, K., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:283809
 This clone was previously sequenced on the 5' end only, this new
 data is from the 3' end
 High quality sequence stop: 451.
 FEATURES
 Source
 1. 346
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:473065"
 /clone_1kb="Soares mouse p3NMF19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCGTAGAGGAGGCGGCGCTTTTCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA was kindly provided by
 Dr. Minoru Ko (Wayne State University)."
 BASE COUNT 134 a 140 c 140 g 132 t
 ORIGIN
 Query Match 6.2% Score 246; DB 9; Length 546;
 Best Local Similarity 70.9%; Pred. No. 1.5e-22;
 Matches 343; Conservative 0; Mismatches 135; Indels 6; Gaps 1;
 Oy 2243 gagatgctacgcttgagctatgatccatccctactacacagatctcttcacgtgt 2302
 |||||
 Db 546 GAGATGCTCAGCTGGGGCGCATCGACTCATCTTACGTAGGCTTACTGCACTGGGT 487
 Oy 2303 tccagtcactgctgagcagctactggaacttcaactgagctgacagctcagcggtgt 2362
 |||||
 Db 486 GCGTGTACAGTACAGGAGATATTGACAGTTTCACAGTGCAGATCAATCAATGGTGA 427
 Oy 2363 ggtgtgtgacgtgaaagtgtgatacagctatcttgataccggtacgtccaaagctgt 2422
 |||||
 Db 426 AGTGTGCTTTGTCAGAGGTGCTGCCCTCTCTGACACAGCAATGCGCTGTGAC 367

Oy 2423 cggacctagcagcagacatctcaacattcagcaagctatgagccacagaccagta 2482
 |||||
 Db 366 GGGGCTGTGATGAGACATCCCAATATTTCAGCAGGTGATGGAGCTGTGCAGAGCCATTA 307
 Oy 2483 cgtgagtttgacatagatgctgacaaaccttaagctacatagctcctacagttgtccttgat 2542
 |||||
 Db 306 TGACCACTGTGACATGACATGCTGAGAGCTGGAGCATATGCTGCCACCGTGTGCTTGTGAGAT 247
 Oy 2543 caacgcaagatgacatcagcccttcgcctatccagcagcagatcaaggtgttc 2602
 |||||
 Db 246 CCATGTTAGGAGTTCACCTGACCTGACCTTATPACCAACAGGATACAGGCTTCTG 187
 Oy 2603 caccatgtagtcacagagtgagaaaccattcccaagaatgagctcttggaagatgttcac 2662
 |||||
 Db 186 CTCAGTGTGCTTTC-----AAGCAGGCTCCACATATGATGCTTGGGATGTTCAT 133
 Oy 2663 tctgtgactactacagcgtctcttgacagggccaacacctgctgagctaaagcaat 2722
 |||||
 Db 132 TCGGAGTTATACAGTGTGTTGACAGAGCCACAAATCGTGTGGCTGGCCAAAGCTAT 73
 Oy 2723 ctga 2726
 |||||
 Db 72 CTGA 69
 RESULT 11
 A1385490 517 bp mRNA linear EST 27-JAN-1999
 LOCUS A1385490
 DEFINITION m181603.y1 Soares mouse p3NMF19.5 mus musculus cDNA clone
 IMAGE:482524 5' similar to TR:Q28950 Q28950 PREPROCHYMOSIN
 PRECURSOR.; mRNA sequence.
 ACCESSION A1385490
 VERSION A1385490.1 GI:4198953
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 517)
 REFERENCE
 AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person
 , B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter
 , E., Kohls, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R. and Wilson, R.
 TITLE The WashU-NCI Mouse EST Project 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:293268
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 possible reversed clone: similarity on wrong strand
 Seq primer: -40RP from Gldco
 High quality sequence stop: 442.
 FEATURES
 source
 1. 517
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:482524"
 /clone_1kb="Soares mouse p3NMF19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',


```
CDS
66..1244
/note="data source:SPTR, source key:P20142, evidence:ISS
homolog to GASTRISIN PRECURSOR (EC 3.4.23.3) (PEPSINOGEN
C)
putative"
/codon_start=1
/protein_id="BAB25990.1"
/db_xref="GI:12843461"
/translation="MKMNVALLCLPLLEAALIRVPLKMKMSIRETMKEOGLKDFLK
NHKYDPGOKHFGKFDYSVLEPMAYMDASVYGEISIGTPPOFVLVLPDSSNLWY
SVYCOSEACTTHRRYRNPSSKSTRYNOGOFPSLOYGFSLGFPGYDLRQSTOVN
OERGLSENEGETNENYVAPDINGMLATPGLSSGGATLALQMLSEGLSOPREYVIG
SQQSGNGQIVFPGVDENLYTGLFTWIPVQELTWITIDFLIGNQASGSSGCG
GIYDTGTSLLVMPAOYLDELQITGAOEGEYFVSCDSVSLPLTFVINGVQFPL
SPSSYIIQEEGSCWVGLESLINAESQPIMLIDVFLRSYAVAFDMGNMNVGLAPSV
"
polyA_signal 1369..1374
/note="putative"
polyA_site 1385
/note="putative"
BASE COUNT 281 a 441 c 336 g 327 t
ORIGIN
Query Match 6.0%, Score 236.2; DB 11; Length 1385;
Best Local Similarity 53.7%; Pred. No. 1.9e-21;
Matches 602; Conservative 0; Mismatches 483; Indels 36; Gaps 4;
QY 1629 gctgagatcacccgcatcctctctcaaaaggttaagtccttcctcgaagcgctgaagaa 1688
DB 111 GCAGCTTGATGAGGCTCCCTCGAAGAAATGAAGAGTATCCGGGAGACATGAAGGA 170
QY 1689 catggaactctagaagactctcttcgaacaacagatgagcatcagacagaactctcc 1748
DB 171 CAAGGTGTCCTCAAGAGACTTTCTGAGAAGAACCAAGATATGACCCCTGGCAGAAATACAC 230
QY 1749 ggcctcgatgaagttgctagcgtgcacttacaa-----ctacccttagtctaa 1799
DB 231 TTTGGCAGTTTGGTGTACTAGAGTGTATGAGCCCATGAGCCCTATATGATGATCTTCC 290
QY 1800 tactttggaagaatctacatcggaaccccgctcaagaagttcacctctctcttatact 1859
DB 231 TACTATGTGTGATGATGACATCGGAGCTCCACCCGAGAACTTCCTGCTTTTCGACACT 350
QY 1860 ggtctcctgactcttggtctccctctatctactgcaagaagcaatgctcgaagaacac 1919
DB 351 GGCTCTCTCAACACTGTGGGTGTCTTGTCTACTGTCAGAGAGGAGCCCTGCACACACAC 410
QY 1920 caaagattcgatcgagaagaagtcgctcaccttcagaacttaggcaaaccttgctata 1979
DB 411 ACCCGCTACAAACCCGACAGGCTCTCCACTACTACACTCAAGGGCAGACCTTCTCCGTG 470
QY 1980 cactcaggttaacagtagatgcaagaatcttagctatgataccgttactgtctccaac 2039
DB 471 CAGTACGGCACCAGGACCTTACCGGCTTTCGGCTATGACACTCTGAGATTCACAAAGC 530
QY 2040 attgtggaacatccaacagacagtagaacttagcaccacaagaacagagtgatgtctcacc 2099
DB 531 ATCCAGGTCCTCAACAGAGAGTTCGGCCTGATGAGATGAGACCTGGCAGCACCAATTTTGC 590
QY 2100 tatgcagaattcgatgcatcctctggtatgcataccacatcgctcgcgtcagaagtaactcg 2159
DB 591 TACGCCCATTTTGAAGGAGATCATGGGCTTACCCGCGCTGTCTTTCACAGGGGGCGCC 650
QY 2160 ataactggtttgaacaatgtagaacccagccttagtagctcaagactgtgttcggt 2219
DB 651 ACCACGCGCTTTCACAGGAGATGTTGGGGAGGGGCGCTGTCTCCACGCCCTCTTCGGTGTGC 710
QY 2220 tacatggaacagaatgagcagaagagcat-----gtcaacgcttggaagctatgacca 2273
DB 711 TACCTTGGCAGCGACAGAGGCTCTAACGGCGGGCAGATTTGTTGGGTGGCGCTGGACGAG 770
QY 2274 tctactaacacagatctcttcaactggtltccagtcactgtgacagcagtaactgcaatlc 2333
```

```
DB 771 AACCTGTACACTGGCGAGACTACCTGATTTCTGTACACCGAGAGCTTTACTGGCAGATC 830
QY 2334 actgtgacagtgatcacatcagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2387
DB 831 ACCATTGACGACTTCTCTATTATGGACACAGAGGCTCTGGCTGTGTCTCTCTCTGTGGCTGC 890
QY 2388 caagctatcttgatbaccggtgacgtlccaaagcttggtcgaacctagcagcagacatctcaac 2447
DB 891 CAAGGCAATTGATGACACAGGACACCTCTGTGCTGCATGAGCCTGCCACAGTCACTGATGAA 950
QY 2448 attcaagcagctatgtgagccacaacagaacagatgacgtgtgtgtgtgtgtgtgtgtgtgt 2507
DB 951 CTTCGACAGACATGAGAGGCCACAGAGAGAGATGATGACATTTTGTACAGCTGCGAC 1010
QY 2508 aacctgactacatgctccacaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2567
DB 1011 AGCGTCACTAGCTGTCTCTTACCTCTACTTTTGTCTCTCAATGCTGTCCAGTTCCCTGTGCA 1070
QY 2568 ccttcgcctatatacagcagcagatcaagggtcttgacacagtgatgattccagag----- 2620
DB 1071 CCGCTTCTCTCATCATCATCAGAGAGAGGCTCTGCTGATGTGGGTGTGAGAGACCTCTCC 1130
QY 2621 -----tgagaaccattcccaagaatgatatcttggaagatgtgttcatctcgtagatc 2672
DB 1131 CTGAACGCTGAGAGTGGCCAGCCCTCTGATCTCTCGGGAGATGCTTCTCTCAGGTTTAC 1190
QY 2673 tacagcgtctttagcagagccacaacactcgltgggcagc 2713
DB 1191 TATGCTGTCTTTCGACATGGCAATMACAGGGTGGGCTTCC 1231
```

RESULT 13

AM868692 619 bp mRNA linear EST 22-MAY-2000
LOCUS M81-SN0064-150500-004-r03 SN0064 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM868692
ACCESSION AM868692
VERSION AM868692.1 GI:8002744
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 619)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
TITLE
JOURN. Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-M81-SN0064-150
500-004-a03&t3=2000-05-15&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 619.
Location/Qualifiers
1..619
/organism="Homo sapiens"
/db_xref="taxon:9606"


```
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="MGI:189566"
/db_xref="taxon:10090"
/clone="2210410L06"
/sex="male"
/tissue_type="stomach"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
66..1256
/note="data source:SPRR, source key:P20142, evidence:ISS
homolog to GASTRICIN PRECURSOR (EC 3.4.23.3) (PERSINOGEN
C)
putative"
/codon_start=-1
/protein_id="BAB25952.1"
/db_xref="GI:12843350"
/translation="MKMYVALLCPLLEALIRVPKKKSTIREMKEGVAKDLK
MKYDQGRKHEKFGDYSLVPPAYMDASTYGEISIGTPPNFLVLEDTGSSNLAY
SSVYQSEACTHTRYNPSKSTYYTQGFSLQYTGSLTGFEDTLVQVNSIQVN
OERGLSENEPNTNFVYAOFDGIMGLAYPGLSGATTAALQMGESALSOPLRGVIG
SOOGNGGOIVFGVDENLYTGLTWIPYQELIWOITIDFLIGNQASGSSGCO
GIYDTGSLIYVMAOYLINELIOTIGAQEEGYGYFVSCDSVSLPILTFVNLGVOPD
SPSSYIIQEBGSCMGLSELINABESQPLMIGDYFLNSYASTMAITGMALPLIS
RDI"
polyA_signal
1368..1373
/note="putative"
polyA_site
1388
/note="putative"
BASE COUNT      282 a      441 c      337 g      328 t
ORIGIN
```

```
Query Match      5.7%; Score 227.4; DB 11; Length 1388;
Best Local Similarity 53.9%; Pred. No. 2,6e-20;
Matches 546; Conservative 0; Mismatches 446; Indels 21; Gaps 3;

QY 1629 gctgagatcacccgcatctctctctacaaagaagtaagtcctcgtlaagcgctgaagaa 1668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111 GCACCTTTGATCAGGTCGCCCGCAAGAAATGAAGATATCCGGAGACCATGAAGAA 170

QY 1689 catggaactctagaagaactctctcagaagaacaacagatgcatcagcagaagaactccc 1748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 171 CAAGGTGTCTCAAGACTTTCTGAAGAACCAAGATATGACCTGTGGCAGAAATATCAC 230

QY 1749 ggtctggtgaagttgctgaagcgtgcacctaccaa-----ctacctgatagtcaa 1799
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 231 TTTCGCAAGTTGGTGTACAGTCTACTCTATGAGCCCATGGCTTATATGATGCTTCC 290

QY 1800 taacttggaaagatctacactcgaaccccgccctcaagaagttcacccgtctcttgatact 1859
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 291 TACTATGTGTAGATCAGATCGGAGCTCCACCCAGAACTTCTGTCCTTTTGACACT 350

QY 1860 ggttcctctgaactctggttcctctctatctactcgaagaacatgctcgaagaacac 1919
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 351 GGCCTCTCAACCTGTGGTGTCTCTGTACTGTCCAGACGAGGCTTGCACACACACAC 410

QY 1920 caaagattcgatccgagaagtcgccaactcagaacttaggcaaaacctgtctata 1979
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 411 ACCGGCTAACACCCCAAGATCTCTCCACTACTACTCACTCAAGGGCAAGACTTTCCTCC 470

QY 1980 caactagatagatgacatgcaagaaactctaggtatgatacgtcatcgtctccaac 2039
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 471 CAGTACGGCACCAGGACCTTACCGGCTTCTTGGCTGTGACACTGTGAGATCCAAAGC 530

QY 2040 atgttgacatcaacagacagtaggacttagcaccacaagaacagtgatgtcttcacc 2099
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 531 ATCCAGGTCCCTTAACAGAGATTGGCTTGAGTGAAGATGAGCTTGCGACCAAAATTTTGC 590

QY 2100 tatcagaattcgatgactcttgatgatacaccatcgctcgctcagatgactcg 2159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 591 TAGCCCAATTGTGACGGGATATGGGCTGGGCTTACCCGGGCTGTCTTTCAGGGGGCGCC 650
```

```
QY 2160 atacctgtgttgacaacatgatagaaccgaacctagtaactaaagactgttctcgtgt 2219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 651 AACACCCCTTTGACAGGGCATGTGGGGAGGGCGCTGTGTCCACCCCTTCTCGGTGC 710

QY 2220 tacatggacag-----gaatgcccagagagacatgtctcagcgtcttgagctatgatcca 2273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 711 TACTTTGGCAGCCAGAGGGGCTCTAACGGCGGGAGATTTGTTCGTGGCGTGGACAGAG 770

QY 2274 tctactacacagatctctcactcagtggttccagtcacgtctgcagcagatctgcaatc 2333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 771 AACCTGTACACTGGCAGATCACCTGATTCGTGTACACCCAGAGAGCTTACTGTGCACATC 830

QY 2334 actgtgacacgtgtcacaccatcagcgtgtgtgtgttcacg-----tgaagtgagatg 2387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 831 ACCATTTGACGACTTCTTATTTGGCAACCCAGGCTCTGTGGTCTCTCTCTGTGGCTGC 890

QY 2388 caactatctcttgatcaccgtgacgtccaaactgtgtcggacctagcagaacatctcaac 2447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 891 CAAGGCATTGTAGACACAGGCACCTCTGTCTGTATGCTGCGCCAGTACCTGAATGAA 950

QY 2448 attcaagaactatcttgagccacacagaaccagatcaggttgatgacatagattcgac 2507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 951 CTTCCTGACAGACATAGAGAGCCAGAGAGAGATGACAGTATTTGTGACCTGCGGAC 1010

QY 2508 aaacttagctatcagctcactacagttgtcttgagatacaaggcaagatgtaaccatgac 2567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1011 AGCGTCAGTACCTCCCTACCTCCTCTTGTCTCATAGGTGTCCAGTTCCTCCCTGCA 1070

QY 2568 cctccgcctctacacagcagaagatcaggtgtgtgcaccagttgattccagag 2620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1071 CCTCTTCTTACATCATTCACAGAGAGGAGGACTCTGTGATGTGGGTCTGAGAGAG 1123
```

```
RESULT 15
AI322423      414 bp  mRNA  linear  EST 23-DEC-1998
LOCUS
DEFINITION m182B01.y1 Soares mouse p3NMFP19.5 Mus musculus cDNA clone
IMAGE:473065 5' similar to TR:Q28075 Q28075 CHYMOSIN C.; mRNA
sequence.
ACCESSION AI322423
VERSION AI322423.1 GI:4056852
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1. (bases 1 to 414)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:283809
This read is a RESSEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
putative full length read
vector to vector length is 655
Seq primer: -40RP from Gibco
High quality sequence stop: 395.
```

FEATURES

location/Qualifiers

I..414

THIS PAGE BLANK (USPTO)

1669 tccgtaagcgctgaagaaacatgactctagaagaactctctcgaagaacaacagatg 1728
160 TGAAGAAAGCGGTGAAGAGACATGGCTTCTGAGAGACTTCTCTGACAAACAGAGATATG 219
1729 gcaatcagcagcaagtaactcggcttggtaagtgtcagcgttgcacttaaccatacc 1788
220 GCATACACACCAAGTACTCCGGCTTCGGGAGGGCCAGGTCCTCCCTGACCAACATACC 279
1789 ttatataataactcttgggaagaatctaccctcggaaccccccacagagttaaccgttc 1848
280 TGGATATCTAGTACTTTGGGAAGATCTACTCGGAGACCCGCCCCAGAGATTTACCGTGC 339
1849 tcttgaactaggttccctcgtactcgtgggttccctctatctactgaagaagcaatgct 1908
340 TGGTTGACACTGGCTCTGACTTCTGGACTTGGATACCTTATCTACTGCAAGAGCAATGGCT 399
1909 gcaagaaccccaagaatttgcaltccgaagaagtcgtccacttcccaaaacttaagcaaac 1968
400 GCAAAAACCAACGCGCTTGACCCGAGAAAGTGTCCACCTTCAGAAACCTGGGCAAGC 459
1969 cctgtctatacactacgtacagtaagtaagcaagaatcttagcctatgataccgtca 2028
460 CCTGTCTATACCACTACGGGAGCAAGCATGCAAGGCGATCTTAGGCTATGACACCGTCA 519
2029 cgtgtccaaacatbtygacatcacaagaacagtaagacttagcaccacaagaacaggtg 2088
520 CTGTCTCAACATTTGTGAGCATCCAGAGACAGTAAAGCCCTGAGACCCAGAGACCGGAG 579
2089 atgtcttccactatgcgaattcgaatcgtatcctcgtgtatggtatccactcgtcgt 2148
580 ACGCTTCACCTATGCGGAATTCACGCGGATCTGGGGATGGCTGACCTCCGCTCGCT 639
2149 cagaactacgatacctgcttbtgacaacatgatgaacccgaactagtagtaact 2208
640 CAGAGTACTGATACCGCTGTTCACAAATGATGAACAGGACACTGTGCGCCCAAGACC 699
2209 tgtctcggattacatgagaagaatggaagaagacatgctcaagcttggagctatg 2268
700 TGTCTCGTTTACATGAGACAGGAATGGCCAGACATCTCACCTGGGGGCCATTCG 759
2269 atccatctactacaagagatctctactcgtgtcgaagctcgtgacagtaagtaagc 2328
760 ACCGCTCTACTACACAGGGGCTCCCTGACTGGGTGCCGCTGACAGTACAGCATGCTGGC 819
2329 aattcactgtgacagtgatcaaccatcagcagtggtgtgtgtgtgtgtgtgtgtgtgt 2388
820 AGTTCACTGTGAGACAGTCAACCATCAAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 879
2389 aagctatcttgatgataccggtacgtcgaagctgtcgaagccttagcagcgaacttcaaca 2448
880 AGGCACTCCCTGAGACACGGGACCTCCCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 939
2449 ttcgaagaagctatgtgagccacacagaacacagtaacggttgaattggaatgattgcgaca 2508
940 TCCAGCAGGCCATTGGAGCCACACAGAACAGTACGATGATTTGACATCAGCTGCGACA 999
2509 accttaagctacatgctcagtgcttcttgagatcaacgaagaatgtaaccacttgacc 2568
1000 ACCTGAGCTACATGCCACAGTGTGTCTTTGAGATCAATGGCAAAATGTAACCACTGACCC 1059
2569 cctcgcgctatcaccagcaggaatcaagggttctgcaccagtgatltccagaatgagaacc 2628
1060 CCTCCGCGCTTACAGGCAAGACAGGCGCTTGTACAGAGGTGTGTGTGTGTGTGTGTGTGT 1119
2629 attccagaagaatgattcttggagatgtgttcatctgtgtgactaataagcttcttgaca 2688
1120 ATTCCAGAAATGATCTGAGGGGATGTTTTCATCCGAGATTAATTAACAGGCTCTTTGACA 1179
2689 gggcacaacacactcgttggctgaactaaagaacttgaagctt 2731
1180 GGGGCAACAACTGTGTGGGGCTGGCCAAAGCATCTGACTCTGT 1222

RESULT 2
US-08-846-021A-6
; Sequence 6, Application US/08846021A
; Patent No. 5948682
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; TITLE OF INVENTION: Preparation of Heterologous proteins on
; TITLE OF INVENTION: Oil Bodies
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,021A
; FILING DATE: April 25, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravalle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 9369-039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 361-1398
; TELEFAX: (416) 361-7311
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 850..1206
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1444..2729
; US-08-846-021A-6
Query Match 22.6%; Score 895.8; DB 2; Length 2733;
Best Local Similarity 88.4%; Pred. No. 3.5e-175;
Matches 972; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
1639 gctaatcaccgcgactctctctacaagtgatgtctccgtaaggcgctgaagaa 1688
1630 GCTAAGATCACACAGATCCCTCTTACAAAGCAAGTCTGTGAGAAAGGCCCTAAGAG 1689
1689 catgactctagaagaactcttgcagaacaacagatagatgcatcagcagaagtaactcc 1748
1690 CATGGGCTTTCGAGAGACTTCTCTCAAGAAACAGCATATGSCATCAGAGCAATACTGCC 1749
1749 ggtctcgttgaagtgtcagcgttgcacttaaccaactcttgatagtaactctggg 1808
1750 GGTCTTCGGGAGGTGGCCAGCGTCCCTGACCAACTACCTGATGATGACTGCTTGGG 1809
1809 aagatctaccctcggaacccgcctcaagagttaaccgttctcttgatactcgttctct 1868
1810 AAGATCTACCTCGGAGCCGCCCCAGAGATTCAACCGTGTGTTGACACTGCTCTCT 1869
1869 gactctcgtgttccctctatctactgcaagaagaatgctgctgaagaaccacaagaattc 1928
1870 GACTTCTGGGATCCCTCTATCTACTGCAAGAGCAATGCTCTCAAAACACACAGCGCTTC 1929
1929 gatcgaagaatgcgtccacttccagaacttgaagaacccttgcctatacactacggt 1988

Db 1930 GACCCGAGAAAGTCCTCCACCTTCAGAACCTGGGCAAGCCCTCTGTATCCACTACGAGG 1989
 1989 acaggtacagcaagagatcttagctatgatataccgtacgtctccaaattgtgac 2048
 1990 ACAGGACACATGACAGGAGCTCTGGGCTATGACACCGTACGTACTCTCCAACTTTGTGGAC 2049
 2049 attcaacagacagtagaacttagcaccacaagacagatgtagtcttccaatgacaa 2108
 2050 ATCCAGACAGACAGTAGGCGCTGAGCAACCAAGAGCCGGGAGAGTCTTACCATCCGCA 2109
 2109 ttcatgtgcatccctgtgtatgacataccatcgctcgctcagagtaactcgaattctgt 2168
 2110 TTCGACGGGATCCTCGGAGATGACCTTACCCCTCGCTCAGAGTACGATGATACCCGCTG 2169
 2169 tttagcaacatgataagacagacagacagtagatgataagactttctcgattacatgac 2228
 2170 TTTGACACATGATGATMAACAGGACCTGGTGGCCCAAGACCTGTCTCGGTTTACATGGAC 2229
 2229 aggaatgagcagagagacatgctcagcgttggagctatgtatcatalcctactacaagaa 2288
 2230 AGGAATGGCCAGAGAGATGCTACGCTGGGGGCCATCGACCCGTCCTACTACACAGGG 2289
 2289 tctcttcaactgtgtccagtaactgtgtcagacagtaactgtgcaattcactgtgacagtgc 2348
 2290 TCCTGTGACTGCTGGGTGCTCCGTGACACTGACAGCTACTGGCACTGCTGAGACAGTGTCT 2349
 2349 accatcaacaggt 2408
 2350 ACCATCAGCGGT 2409
 2409 acgtccaaagctgtgtcagac 2468
 2410 ACCTCAAGCTGT 2469
 2469 acacagaacacagtaactgt 2528
 2470 ACACAGAACCTGATGAGT 2529
 2529 gt 2588
 2530 GTGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2589
 2589 gatcaagaggtgtgtcagac 2648
 2590 GACCAAGGCTTCTGTACAGTGGCTTCCAGAGTGAATAATCCACAGAAATGATCTCG 2649
 2649 ggaagatgtgttcatctgt 2708
 2650 GGGGATGTTTTCATCTCGAGAGTATACAGCGCTTTGACAGGGCAACAACCTCGTGGGG 2709
 2709 ctgactaagaatctgaa 2727
 2710 CTGGCCAAAGCCATCTGAA 2728

Query Match 22.4%; Score 884.6; DB 6; Length 2732;
 Best Local Similarity 82.6%; Pred. No. 6.9e-173;
 Matches 1024; Conservative 0; Mismatches 214; Indels 1; Gaps 1;
 1610 cgtgtgttactacagcgt 1669
 669 cgtgtgttactacatgt 728
 1670 ccgttaagcgtgtgaaggaacatgtgactctagaagaactcttgcagaagaacatgtgtgtgt 1729
 729 gaggaaagcgtgtgaaggaacatgtgactctagaagaactcttgcagaagaacatgtgtgtgt 788
 1730 catcagaacagatgt 1789
 789 catcagaacagatgt 848
 1790 tgatgtcaatacttctgt 1849
 849 ggaacagtaacttctgt 908
 1850 cttgtatgt 1909
 909 gttgtacactgt 968
 1910 caaagaacacacaaagatgt 1969
 969 caaagaacacacaaagatgt 1028
 1970 ctgtgttatacactcgt 2029
 1029 cctgtgtatacactcgt 1088
 2030 tgtgttccaaactgt 2089
 1089 tgtgttccaaactgt 1148
 2090 tgtgttccaaactgt 2149
 1149 cgtgttccaaactgt 1208
 2150 aagatgt 2209
 1209 aagatgt 1268
 2210 gttctcgt 2268
 1269 gttctcgt 1328
 2269 atccatcctactacacagatgt 2328
 1329 accgtgtcctactacacagatgt 1388
 2329 aattcactgt 2388
 1389 agttcactgt 1448
 2389 aagatgt 2448
 1449 aggcatacctgt 1508
 2449 ttcagaacagatgt 2508
 1509 ttcagaacagatgt 1568
 2509 acctagatgt 2568
 1569 acctagatgt 1628
 2569 cctcgcctatacagcagacagatgt 2628
 1629 cctcgcctatacagcagacagatgt 1688

RESULT 3
 5217891-14
 : Patent No. 5217891
 : APPLICANT: BRAKE, ANTHONY J.; VAN DEN BERG, JOHAN A.
 : TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYVEROMYCES
 : A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS
 : POLYPEPTIDES
 : NUMBER OF SEQUENCES: 23
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/507,398
 : FILING DATE: 09-APR-1990
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 78,551
 : FILING DATE: 28-JUL-1987
 : SEQ ID NO:14:
 : LENGTH: 2732
 : 5217891-14

```

Oy 2629 attcccaaaatgagctcttgggaagatgtgtcattccgtgagactacaagcgctttgaca 2688
Db 1689 attcccaaaaatgagctcccgggggagatgttccatccggaagataacacgctttgaca 1748
Oy 2689 gggccaacaacctcgttggctgtagctaaagcaatcgtgaagcttaataagtaagactaaa 2748
Db 1749 gggccaacaacctcgttggcgctggcgcaaacgcatcgtatccgcagcttggtaacaagct 1808
Oy 2749 atgcattagatgttaagaagctcattagagagacatgtaatatgtatccgacacgttaacg 2808
Db 1809 gccaaagcttaagtaattactttaaagcttgcatttaataaatttctttatg 1868
Oy 2809 tataatactagagctccatccactcttctatgataa 2847
Db 1869 cttaatgactaagttcaattatataataattatg 1907

RESULT 4
US-08-318-193-15
; Sequence 15, Application US/08318193
; Patent No. 5641663
; GENERAL INFORMATION:
; APPLICANT: GARVIN, Robert T.
; APPLICANT: MALEK, Lawrence T.
; TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION
; TITLE OF INVENTION: OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
; TITLE OF INVENTION: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
; TITLE OF INVENTION: PROTEINS FROM STREPTOMICES
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318,193
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,314
; FILING DATE:
; APPLICATION NUMBER: US 07/224,568
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/116 CACO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid:
; DESCRIPTION: Synthetic DNA oligonucleotide
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 399..830
; US-08-318-193-15

```

[illegible]

```
RESULT 6
5200327-5
: Patent No. 5200327
: APPLICANT: GARVIN, ROBERT T., MALEK, LAWRENCE T.
: TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF
: BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING
: FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM
: STREPTOMYCES
: NUMBER OF SEQUENCES: 24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/224,568
: FILING DATE: 26-JUL-1988
: SEQ ID NO: 5:
: LENGTH: 846
: 5200327-5

Query Match          5.0%; Score 199.8; DB 6; Length 846;
Best Local Similarity 76.2%; Pred. No. 1.8e-32;
Matches 246; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1627 ctgctgagataccgcgcatctctctacaagaagtaagtcctcgtgaagcgctgaag 1686
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 ctgcagagactactagaatccattgtacaaggtgaagcttcttggaagcgctgaag 576

QY 1687 aacatgagactctagaagactctctgcagaacaacagatagtcacagcaagtaact 1746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 aacagcggttctggaagactctctgcagaacaacagatagtcacagtaact 636

QY 1747 ccggtctcggtgaagctctctagcgtgcacactaccactactatagtaactcttg 1806
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 ctggttcgtggaagctcgtctccgtctccatctgaccactactctggaatcccaatcttg 696

QY 1807 ggaagatctactcctggaaaccccgctcaagaagttaccggttctcttgatactgttct 1866
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 gtaagatctactagtaagtaaccacacaaagaaatctactgtctgttcgacacggttct 756

QY 1867 ctgactctcggttcctccctctactctcgaagagcaatgctgcgaagaaccccaagaat 1926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 ctgactctcggttcctccctctactctcgaagagcaatgctgcgaagaaccccaagaat 816

QY 1927 tcgaltcgaagaagtcgtccacc 1949
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 tcgaccgaagaagagctctagc 839

RESULT 7
US-08-318-193-13
: Sequence 13, Application US/08318193
: Patent No. 5641663
: GENERAL INFORMATION:
: APPLICANT: GARVIN, Robert T.
: APPLICANT: MALEK, Lawrence T.
: TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION
: OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
: PROTEINS FROM STREPTOMYCES
: NUMBER OF SEQUENCES: 91
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/318,193
```

```

: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,314
: FILING DATE:
: APPLICATION NUMBER: US 07/224,568
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 18740/116 CACO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-9300
: TELEFAX: (703) 683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 838 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: Other nucleic acid;
: DESCRIPTION: Synthetic DNA oligonucleotide
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 399..836
: US-08-318-193-13

Query Match          5.0%; Score 199.4; DB 1; Length 838;
Best Local Similarity 76.3%; Pred. No. 2.1e-32;
Matches 245; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1627 ctgctgagataccgcgcatctctctacaagaagtaagtcctcgtgaagcgctgaag 1686
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 ctgcagagactactagaatccattgtacaaggtgaagcttcttggaagcgctgaag 576

QY 1687 aacatgagactctagaagactctctgcagaacaacagatagtcacagcaagtaact 1746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 AACACGGTTGTTGGAAGACTCTTGCAAAACCAACAAATACGGTATCTCTCCAACTACT 636

QY 1747 ccggtctcggtgaagctctctagcgtgcacactaccactactatagtaactcttg 1806
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 CTGTTTCGGTGAAGTCGCTTCCTCCATGTGACCAACTACTTGACATCCCAATCTTGG 696

QY 1807 ggaagatctactcctggaaaccccgctcaagaagttaccggttctcttgatactgttct 1866
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 GTAAGATCTACTTAGTAACTACCCACCAAGAAATTCATCTGTTGCGACACCGGTTCT 756

QY 1867 ctgactctcggttcctccctctactctcgaagagcaatgctgcgaagaaccccaagaat 1926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 CTGACTTCTGGGTCCCATCGATTCTGTAGTCCACACGCTTGTAAGAACCAACCAAGAT 816

QY 1927 tcgaltcgaagaagtcgtccaca 1947
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 TCGACCCCAAGAAAGAGCTCTA 837

RESULT 8
US-09-032-523-6
: Sequence 6, Application US/09032523
: Patent No. 6232454
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Corley, Neil C.
: APPLICANT: Guegler, Karl
: APPLICANT: Baugh, Mariah
: TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
```

```

STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0479 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PANTU01
CLONE: 151565
US-09-032-523-6

```

```

Query Match          4.9%  Score 192.2; DB 4; Length 2073;
Best Local Similarity 56.5%  Pred. No. 8.3e-31;
Matches 406; Conservative 0; Mismatches 296; Indels 15; Gaps 2;

```

```

OY 1640 ccgactctctctcaagaagtaagtcctccgtaagcgctgaaggaacatgacttct 1699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 104 CAGGTTGCCCTCAGAGAGCATCCGTCCTCAAGAGAAGCTGGCGGACGAGCAGCT 163
OY 1700 agaagactctctcgaaacaacaglatgcaatca-----gcagaagctactccgctt 1753
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 164 CTCGTAGTTCTGGAATCCATTAATTTGGACATGCCAGTTCACCGAGTCTGTCAT 223
OY 1754 cgtgaagtctgtagcggtccactaccactactgatactacttgggaagat 1813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 224 GACCCAGAGTCCAAAGAACCCCTCATCACTACTTGGATATGAAATCTTGGCACTAT 283
OY 1814 ctactcggaaccgctcaagaagctacacgtctctcttgaactgttctctgaact 1873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 284 CTCCTATTGGCTCCACACACAGAACTTCACTCTCATCTTGCAGACTGGCTCTCAACT 343
OY 1874 ctgggtctccctatctactcaagaagcaatgctcgaaagaacacccaagaattcgatcc 1933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 344 CTGGGTCCTCTGTGTACTGTGACTGACCTGACAGCCCTGCAAGACGACAGAGTTCACGCC 403
OY 1934 gagaagaagtcgtccactctcagaacttaggcaaacctgtctatacatcaggtacag 1993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 404 TTCCAGTCCACACATACAGCCAGCAGTCATCTTCTCCATTCAGTATGGAACCGG 463
OY 1994 tagcatcaaggactataggctatgataccgtcaactgtctcaaatgtgtgacattca 2053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 464 GAGCTTTGTCGGGATCATTTGAGCCGACCAAGTCTCTGTGGAAGACATACCGTGGTTGG 523
OY 2054 acagacagtaggacttagcaccacaagaacaggtgagtctcaacctatgcagaattcga 2113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 524 CCAGCAGTTTGGAGAAAGGTCAACAGAGCCAGCCAGACCTTTGTGATGCAAGATTGA 583
OY 2114 ttgactcttgtagtgaataccacatcgctcggtcaaggtactgatatcctggtgttga 2173
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 584 TGAATTCCTGGGCTGGGATACCCCTTCTTGCTGTGGAGAGATGATCCAGTATTGA 643
OY 2174 caaatatgaaccgacacttagtagctcaagacttgctcgtttacatgacagaa 2233
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 644 CAACATATGCTCAGAACCTGGTGACTTGGCCATGTTTCTCTCATGACATTA 703
OY 2234 -----tgccagagagacatgctcaccgcttgagactatgatacactactaac 2284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 704 CCCAGAGGTGTGTCGCCGAGACGAGCTGATTTTGGAGGGTACGACACCATTCCTC 763
OY 2285 agaatcttcaactggttccagtcactgtgcagcagtagtgcgaattcactgtgaca 2343
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 764 TGGAGCCTGAATTGGTCCAGTACCAAGCAAGCTTACTGCAAGATTGCATGATA 822

```

```

RESULT 9
US-08-723-938-4
Sequence 4, Application US/08723938
Patent No. 5776759

```

```

GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/723,938
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0125 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

```

```

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
LIBRARY: LUNGNOT02
CLONE: 312099
US-08-723-938-4

```

```

Query Match          3.7%  Score 146.6; DB 1; Length 1299;
Best Local Similarity 50.4%  Pred. No. 1.7e-21;
Matches 421; Conservative 0; Mismatches 399; Indels 15; Gaps 2;
OY 1769 cgtgcacttaccactacttgatgataactcttgggaagatctactcgaacccc 1828

```

```

Db 217 CGTACCTCTCGAACTACAGAGATGTCAGATATTTGGGAAATTTGGGAGAACCC 276
1829 gcccaagaagtcacccgtctccttgatactggtccctcgtactcgtggtccctccta- 1887
Db 277 TCCACAAAACCTTCACTTGCTGGTACACCTGCTCCCAATCTCTGGGTCCCGCAG 336
1888 -----tctactgcaagaacatgctcgtcaagaacccaagaattcgatccgaagaagtc 1942
Db 337 GAGATGCCACTTCTTCACTTGCTGGTACACCAAGATTTGATCCCAAGCCCTC 396
1943 gtccacttcgaagaactaggcaaaccttgcctatacactaaggtacaggtacagtcga 2002
Db 397 TAGCTCTTCCAGGCCAATGAGGACCAAGTTGCCATTCATATGAACTGGCGGCTAGA 456
2003 agaatcttagctatgatacgttaccgttactgctcgaacattgtgacatccaagaacgt 2062
Db 457 TGGAACTCTGAGCGAGCAAGCTGCAATTTGGTAATCAAGGGTGCATCAGTATTTT 516
2063 aggaacttagcaaccaagaacaggtgatagtcttcacactatgcaagaattcgatcct 2122
Db 517 CGGGAGGCTCTCTGGGAGCCACGCTGCTCTGCTTTGCCATTTTGATGGATATT 576
2123 tggatagcataccatcgctcgctgcagagtaactgatacctgtgttgacaacatgat 2182
Db 577 GGGCTCGGTTTCCATTTCTGTGTGAAGAGATTGGCCCGCATGATGATGATGAT 636
2183 gaacgacaacctagtagtcaagaacttgtctggttacaatgacaagaatgagcaaga 2242
Db 637 GGAGCAGGGGCTATTGATGAAGCTGCTGCTCTTCTTACCTCAACAGGAGCCTGAAGA 696
2243 g-----agcatgctcaacgttgagctatgatacctactcactacagaagatcct 2293
697 GCCGTGATGAGAGAGAGCTGGTCTCGGGGGCTCGGAGCCCGGACACTATCCACCCCT 756
2294 tcactgggtccagtcactgtgcagcagtaactggaactcactgtggaacagtgtaacat 2353
Db 757 CACCTTCTGTCAGTCACAGGTCCCGCTACTGCGAGATTCACATGAGAGCGTGAAGGT 816
2354 caggggtgtgtgtgtgtgataagtgatgtaagctcttgatccgtgtaacagtgatc 2413
Db 817 GGGGCCAGGGCTGACTCTGTGGCCAGGGCTGTCTCCATCTCGATACGGGAGCTGC 876
2414 caagctgtgctgacacgaacgaacatctcaacatcagaacgaactatgtagagcacaca 2473
Db 877 CCTCATCAGAGACCCATGAGGAGATCCGGGCTCGATGACGACATTTGGGGGAAATGCC 936
2474 gaacagtaaggtgagttgacatagattgcaaaccttagctacatgctacagttgt 2533
Db 937 CTTCGTGGCTGGGAGTACATCATCTGTGTCGGAATCCCAAGACTCCCGCAGTCTC 996
2534 ctgtgagtaacaggaagatgtaacccatgaacccctcgctctatacagcag 2588
Db 997 CTTCCTTCTGGGGGGTCTGTGTTAACTTCAGGCCCATGATGATGATCAG 1051

```

RESULT 10
 US-09-080-538-4
 : Sequence 4, Application US/09080538
 : Patent No. 5965129

: GENERAL INFORMATION:
 : APPLICANT: Bandman, Olga
 : APPLICANT: Coleman, Roger
 : TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Incyte Pharmaceuticals, Inc.
 : STREET: 3174 Porter Drive
 : CITY: Palo Alto
 : STATE: CA
 : COUNTRY: U.S.
 : ZIP: 94304

COMPUTER READABLE FORM:

```

: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/080,538
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/723,938
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0125 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1299 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
: IMMEDIATE SOURCE:
: LIBRARY: LUNGNOT02
: CLONE: 312099
: US-09-080-538-4

```

Query Match 3.7%, Score 146.6; DB 2; Length 1299;
 Best Local Similarity 50.4%; Pred. No. 1.7e-21;
 Matches 421; Conservative 0; Mismatches 399; Indels 15; Gaps 2;

```

QY 1769 cgtgcacttaccacactccttgatagtcataactcttggaaagatcctcctggaaccc 1828
Db 217 CGTACCTCTCGAACTACAGAGATGTCAGATATTTGGGAAATTTGGGAGAACCC 276
1829 gcccaagaagtcacccgtctccttgatactggtccctcgtactcgtggtccctccta- 1887
Db 277 TCCACAAAACCTTCACTTGCTGGTACACCTGCTCCCAATCTCTGGGTCCCGCAG 336
1888 -----tctactgcaagaacatgctcgtcaagaacccaagaattcgatccgaagaagtc 1942
Db 337 GAGATGCCACTTCTTCACTTGCTGGTACACCAAGATTTGATCCCAAGCCCTC 396
1943 gtccacttcgaagaactaggcaaaccttgcctatacactaaggtacaggtacagtcga 2002
Db 397 TAGCTCTTCCAGGCCAATGAGGACCAAGTTGCCATTCATATGAACTGGCGGCTAGA 456
2003 aggaacttaggctatgatacgttactgctcgaacattgtggaacattggaacagagag 2062
Db 457 TGGAACTCTGAGCGAGCAAGCTGCAATTTGGTAATCAAGGGTGCATCAGTATTTT 516
2063 aggaacttagcaaccaagaacaggtgatagtcttcacactatgcaagaattcgatgcatcct 2122
Db 517 CGGGAGGCTCTCTGGGAGCCACGCTGCTCTGCTTTGCCATTTTGATGGGATATT 576
2123 tggatagcataccatcgctcgctgcagagtaactgatacctgtgttgacaacatgat 2182
Db 577 GGGCTCGGTTTCCATTTCTGTGTGAAGAGATTGGCCCGCATGAGATGATGATGAT 636
2183 gaacgacaacctagtagtcaagaacttgtctggttacaatgacaagaatgagcaaga 2242
Db 637 GGAGCAGGGGCTATTGATGAAGCTGCTCTTCTTACCTCAACAGGAGCCTGAAGA 696
2243 g-----agcatgctcaacgttgagctatgatacctactcactacagaagatcct 2293

```



```

:
: TOPOLOGY: Linear
: MOLECULE TYPE: Genomic DNA
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: No. 5968816 applicable
: ORIGINAL SOURCE:
: ORGANISM: homo sapiens
: STRAIN: not applicable
: INDIVIDUAL ISOLATE: not applicable
: DEVELOPMENTAL STAGE: not applicable
: HAPLOTYPE: not applicable
: TISSUE TYPE: blood
: CELL TYPE: Leucocyte
: CELL LINE: HeLa
: ORGANELLE: not applicable
: IMMEDIATE SOURCE:
: LIBRARY: not applicable
: CLONE: not applicable
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT: not applicable
: MAP POSITION: not applicable
: UNITS: not applicable
: FEATURE:
: NAME/KEY: This is the DNA sequence
: NAME/KEY: claimed in 15(v1) as the Cathepsin gene in FIG. 15.
: LOCATION: not available
: IDENTIFICATION METHOD: experiment-
: IDENTIFICATION METHOD: in specification
: OTHER INFORMATION: prevention of IFN-2
: OTHER INFORMATION: promoted cell death
: PUBLICATION INFORMATION: not available
: US-08-631-097-7

```

```

Query Match 3.5%; Score 138.2; DB 2; Length 2038;
Best Local Similarity 51.0%; Pred. No. 1e-19;
Matches 443; Conservative 0; Mismatches 378; Indels 48; Gaps 3;

```

```

QY 1769 cgtgcacttaccacttactgtatgatacttgggaagatctactcggaaacccc 1828
DB 255 CGAGGTCCTCAAGAACTACATGAGCCCGCAGTACTGCGGAGGAGATGGCAGCGCC 314
QY 1829 gectaaagatcaccgctctcttggatactgttctccttgaacttctgttccctcat 1888
DB 315 CCCCCAGTCTTCAAGTGTCTTCGACACGGGCTCTCAACCTGTGGTCCCTCCAT 374
QY 1889 ctactgcaaa-----gagcaatgctcgaagaacacccaagaatcgcagaaagtc 1942
DB 375 CCACCTGCAAACTGCTGGACATGCTTGTGATCCACCAAGTACAAACGACAAAGTC 434
QY 1943 gtccacctccagaacttaggcaaaccttgcctatacactacggtacagtgatgca 2002
DB 435 CAGCACTCACTGTAAGAAAGTACCTCGTTTGACATCCACTATGGCTCGGCGCCTTC 494
QY 2003 aggaatcttaggctatgatacgttcaactgtctccaaacttggagatctcaa----- 2054
DB 495 CGGTAACCTGAGCAGACACTGTGCGGCTGCCCTCCAGTCAAGCTCGTCAAGCTCTGC 554
QY 2055 -----cagacagtaggacttaggcaaaccttgcctcgaagaacagtgta 2089
DB 555 CCTGGGCGGTGCAAAAGTGGAGAGGCTCTTTGGGGAGGCGCCCAACGACGACGAT 614
QY 2090 tctctcaactatgcagaatctgatgcaactcttgatgatacaccatcgctcgcgtc 2149
DB 615 CACCTTCATCGCAGCAAGTATGATGATCCCTGGGCGCATGGGCTTACCCCGCATCTCCG 674
QY 2150 agaatcttaggactgttcttgaacaatgatacgaacacactgtatgcttaagact 2209
DB 675 CAACAACGTGCTGCCGCTTCGACAACTGATGACAGCAAGAGCTGTGGACGACAACT 734
QY 2210 gtctcgtgttcaatgagaaagaa-----tgccagagagagatctcagcgtcttg 2260
DB 735 CTCTCTCTTCTACTGAGGAGGAGCCAGATGCGCAGCCTGGGGGTGAGCTGATGCTGG 794

```

```

QY 2261 agctatgatccatctactacacagagatctcttacttgggttccagtcactgtgcaga 2320
DB 795 TGGACACAGCTCCAGTATTTACAAAGGTTCTGTCTACTCATGATGATCACCAGCAAGC 854
QY 2321 gtaactggcaattcactgttggagaagatgtaaccatcagcgtgtgtgttgcatagtgaag 2380
DB 855 CTACGTGGCAGGTCCACCTGAGCAGCAAGTGAAGTGGCCAGCGGCTGACCTTGCAGAGA 914
QY 2381 tgaatgcaagctatcttggatacgggtacggtcacaagctgtgtcggacttagcaagacat 2440
DB 915 GGGCTGTGAGGCGCATTTGTGACACAGGCACTTCCCTCATGTGTGGGCGCGGTGATGAGGT 974
QY 2441 tctcaacattcagcaagctatgtgagccacacagaacaagtaagtgatgttgcataga 2500
DB 975 GCGCAGGCTGAGAGGCGCATGCGGCGCGCTGATTCAGGCGGATCATGATGCC 1034
QY 2501 ttgcgacaaccttagctatagctcactcagttgtcttggatcaacgcaagatgtacc 2560
DB 1035 CTGTGGAAGGTGTCCACCCTGCCCGCATTCACACTGAAGCTGGAGGCAAGGCTACAA 1094
QY 2561 actgacccccctcgcctataccagcagg 2589
DB 1095 GCTGTCCCGAGAGACTACACCTCAAG 1123

```

```

RESULT 14
US-08-810-712-11
: Sequence 11, Application US/08810712G
: Patent No. 6160106
: GENERAL INFORMATION:
: APPLICANT: Yeda Research and Development Co. LTD
: TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and
: TITLE OF INVENTION: Use of said Genes and Proteins
: FILE REFERENCE: sequence list
: CURRENT APPLICATION NUMBER: US/08/810,712G
: CURRENT FILING DATE: 1997-03-03
: EARLIER APPLICATION NUMBER: PCT/US94/11598
: EARLIER FILING DATE: 1994-10-12
: NUMBER OF SEQ ID NOS: 31
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 11
: LENGTH: 2038
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-08-810-712-11

```

```

Query Match 3.5%; Score 138.2; DB 4; Length 2038;
Best Local Similarity 51.0%; Pred. No. 1e-19;
Matches 443; Conservative 0; Mismatches 378; Indels 48; Gaps 3;

```

```

QY 1769 cgtgcacttaccacttactgtatgatacttgggaagatctactcggaaacccc 1828
DB 255 CGAGTGTCTCAAGAACTACATGAGCCCGCAGTACTGCGGAGGAGATGGCAGCGCC 314
QY 1829 gectaaagatcaccgctctcttggatactgttctccttgaacttctgttccctcat 1888
DB 315 CCCCCAGTCTTCAAGTGTCTTCGACACGGGCTCTCAACCTGTGGTCCCTCCAT 374
QY 1889 ctactgcaaa-----gagcaatgctcgaagaacacccaagaatcgcagaaagtc 1942
DB 375 CCACCTGCAAACTGCTGGACATGCTTGTGATCCACCAAGTACAAACGACAAAGTC 434
QY 1943 gtccacctccagaacttaggcaaaccttgcctatacactacggtacagtgatgca 2002
DB 435 CAGCACTCACTGTAAGAAAGTACCTCGTTTGACATCCACTATGGCTCGGCGCCTTC 494
QY 2003 aggaatcttaggctatgatacgttcaactgtctccaaacttggagatctcaa----- 2054
DB 495 CGGTAACCTGAGCAGACACTGTGCGGCTGCCCTCCAGTCAAGCTCGTCAAGCTCTGC 554
QY 2055 -----cagacagtaggacttaggcaaaccttgcctcgaagaacagtgta 2089

```



```
Db 555 cctggcggtgtcaagtgagagagcaagctcttgggagccaccacagcagcgcat 614
QY 2090 tgtcttacccttgcagaatctcatgagctcttggatgcatcaccatcgctgcgc 2149
Db 615 cacccttcacgagcagaatctcgatggcatcctggcagatgaccccgacatccgct 674
QY 2150 agagtaactcgatcctgtgtttgacaacatgaagaaccacactagtagtcaagact 2209
Db 675 caaacacgtgcgcgcttcctcgacaactgacgacgaagaagctggtgaccagaact 734
QY 2210 gtcttcggtttacatgacagaa-----tgccagagagacatgctcaagcttg 2260
Db 735 ctctccctctcaccgagacagagaccagatgacgacgctgggggtgagctgatgctgg 794
QY 2261 agcttatgcatccctactacagagatcttcaactggttcaagctggtgacga 2320
Db 795 tggcacaagatccaagatattacaagggtctctgtctcctacccaatgtaaccgcaagc 854
QY 2321 gtactggaattcaactgttggacagtgacacacacacagcgggtgtgtgtgtgacatgaaag 2380
Db 855 ctactggaaggttcacacccctggaccaggttgaggttgccacagcggcctgacccctgtgcaaga 914
QY 2381 tggatgtcaagctatcttggataccggtacgttccaaagctgtgcggaacctagcagcgacat 2440
Db 915 gggcgtgtgagcattgtggacacagcacttccctcatgtgtggtggtggtggtggtggtggt 974
QY 2441 tctcaaatctcagcaagctattggagggccacagacacacagcgttcgtgtgattgacataga 2500
Db 975 ggcgaagctgcagaaagccatcgcggcgctgcgctgattcagcgagatgacatgattcc 1034
QY 2501 ttgcagacaaccttgaatcatatccttaccagttgtcttggatcaacagcaagatgtaccc 2560
Db 1035 ctgtgagagaggtgtcacccttgcgcgcgacatcaactgaaagctggagagaaagctacaa 1094
QY 2561 actgaacccctcgcctatataccagccagg 2589
Db 1095 gctgtcccccagagactacacgctcagg 1123

RESULT 15
PCT-US92-08090-1
: Sequence 1, Application PC/TUS9208090
: GENERAL INFORMATION:
: APPLICANT: Tang, J. N.
: TITLE OF INVENTION: Fusion Protein Genes for Treatment of
: TITLE OF INVENTION: Aids
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kilpatrick & Cody
: STREET: 100 Peachtree Street
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: U.S.
: ZIP: 30303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/08090
: FILING DATE: 19920922
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Padst, Patrea L.
: REGISTRATION NUMBER: 31,284
: REFERENCE/DOCKET NUMBER: OMR129
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404-572-6555
: TELEFAX: 404-572-6555
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
```

```
: LENGTH: 2465 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: Single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapien
: TISSUE TYPE: Epithelial
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 1..6
: OTHER INFORMATION: /note= "Restriction site"
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 2460..2465
: OTHER INFORMATION: /note= "Restriction site"
: PCT-US92-08090-1
```

```
Query Match 3.5%; Score 138.2; DB 5; Length 2465;
Best Local Similarity 51.0%; Pred. No. 1,1e-19;
Matches 443; Conservative 0; Mismatches 378; Indels 48; Gaps 3;
```

```
QY 1769 cgtgcacacttaccacttacttgaatgaatcatttggagaatcttacttgaacccc 1828
Db 1424 CAGAGTGCCTCAAGAACTACATGACGCCCACTACTACGGGGAATTTGGCATGGGACGCC 1483
QY 1829 gctcaagaagctcacaccgttctcttgaatcgtgttctcctgaacttctggttccctcat 1888
Db 1484 CCCCAGTCTCTCAAGTGTCTTGACACACGGGCTCTTCAACCTGTGGGTCCCTCCAT 1543
QY 1889 ctactgca-----gagcaatgcttgaagaacacacaaagattcgatccgaagaatc 1942
Db 1544 CCACGTCAAACTGCTGACATCGCTTGCTGATCCACCAAGTACAAACGACGACAAATC 1603
QY 1943 gtccacacttcagaacttggcgaacccctgtcttaccacttcgttaccagtgatga 2002
Db 1604 CAGCACTACGTGAAAGATGATGTAACCTCGTTTGACATCCATATGGCTCGGAGCCTCTC 1663
QY 2003 aggaatcttaagctatgataccgtcactgtctccaacatttggacatcaa----- 2054
Db 1664 CCGGTACTGTGACCGACGACACTGTGTGGGTGCCGCGCAGTCAAGGCTGTCAAGCCTTCC 1723
QY 2055 -----cagacagtaggaacttagcaccacaaagacagtgta 2089
Db 1724 CCTGGCGGTGTCAAAAGTGAGAGGCAAGGTCTTTGGGGAGGCCACCAAGCAGGCAT 1783
QY 2090 tgtcttacccttgcagaatctgagatccttcttgaatgataccatcgctgcgc 2149
Db 1784 CACCTTCATCGCAGCCAAAGTTGATGGCATCTTGCGCATGGCTTACCCCGCATCTCCGT 1843
QY 2150 agagtaactcgatcctgtgtttgacaacatgaagaaccacactagtagtcaagact 2209
Db 1844 CAACAACGTGCTGCCCGCTTGTGACAACTGATGACGACGAGACTGTTGACAGAACT 1903
QY 2210 gtcttcggtttacatgacagaa-----tgccagagagacatgctcaagcttg 2260
Db 1904 CTTTCCTCTTCTACTGAGAGGAGGACCCAGATCGCAGCTGTGGGGTGAAGTATCTGTGG 1963
QY 2261 agcttatgcatccctactacagagatcttcaactggttcaagctggtgacga 2320
Db 1964 TGGCACAAGCTCAAGTATTACAAAGGTTCTGTCTACTCAAGTAATGTACCCGCAAGGC 2023
QY 2321 gtactggaattcaactgttggacagtgacacacacacagcgggtgtgtgtgtgacatgaaag 2380
Db 2024 CTACTGCAAGGTCCACCTGAGACCAAGTGAAGTGGCCAGCGGCTGACCTGTGCAAGA 2083
QY 2381 tggatgtcaagctatcttggataccggttcaagctgtgtcgaacctagcagcgacat 2440
Db 2084 GGGCTGTAGGCCAATTGTGACACAGGCACTTCCTCATGTGTGGGCCCGGTGATGAGGT 2143
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 21:01:50 ; Search time 9238.6 Seconds
(Without alignments)
2656.986 Million cell updates/sec

Title: US-09-643-755B-1
Perfect score: 1173
Sequence: 1 atgaacttccttaagctctt.....gactagctaaagcaatctga 1173

Scoring table:
OLIGO_NUC
Gapop-60.0-, Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	1173	100.0	1173	6	AX088019	AX088019 Sequence
2	1173	100.0	3957	6	AX088021	AX088021 Sequence
3	56	4.8	1269	4	BOVCHYMOC	J00004 Bovine chym
4	49	4.2	1098	6	E00079	E00079 CDNA sequen
5	49	4.2	1098	6	E02341	E02341 CDNA sequen
6	49	4.2	1101	4	AF177290	AF177290 Bubalus a
7	49	4.2	1115	12	SYNPROCA	M22593 Synthetic b
8	49	4.2	1175	6	I04058	I04058 Sequence 5
9	49	4.2	1240	6	AR002347	AR002347 Sequence
10	49	4.2	1275	4	BOVCHYMOC	J00002 bovine chym
11	49	4.2	1289	6	E00144	E00144 CDNA encod1
12	49	4.2	1291	6	A15836	A15836 chymosin ge
13	49	4.2	1305	4	BOVCHYMOB	J00003 bovine chym
14	49	4.2	1460	6	E00042	E00042 DNA coding
15	49	4.2	1460	6	E00295	E00295 CDNA encod1
16	49	4.2	2726	6	I08097	I08097 Sequence 5
17	49	4.2	2733	6	AR073077	AR073077 Sequence
18	49	4.2	2962	6	I08098	I08098 Sequence 1
19	44	3.8	1094	4	BP019786	U19786 Bos primlge
20	38	3.2	1290	6	E00075	E00075 CDNA encod1
21	38	3.2	1291	6	A15633	A15633 preprochymo
22	38	3.2	1311	6	E00108	E00108 DNA coding
23	37	3.2	119	4	BOVCHYMOC	M14076 Bovine chym
24	35	3.0	171	4	BOVCHYMOC	M14070 Bovine chym
25	35	3.0	637	6	E05472	E05472 DNA sequenc
26	33	2.8	179	6	I04059	I04059 Sequence 6
27	33	2.8	1172	4	CDR131677	M131677 Camelus d
28	32	2.7	1335	4	BOVCHYMOC	M14077 Bovine chym
29	32	2.7	1292	4	OAPECHY	X53037 Lamb mRNA f
30	30	2.6	134	4	BOVCHYMOC	M14074 Bovine chym
31	30	2.6	165	4	BOVCHYMOC	M14075 Bovine chym
32	29	2.5	1460	6	E00042	E00042 DNA coding
33	29	2.5	1460	6	E00295	E00295 CDNA encod1
34	27	2.3	139	4	BOVCHYMOC	M14072 Bovine chym
35	25	2.1	138	4	BOVCHYMOC	M14071 Bovine chym
36	23	2.0	1140	6	AX319837	AX319837 Sequence
37	23	2.0	1722	8	AT051036	U51036 Arabidopsis
38	23	2.0	1820	8	AY056403	AY056403 Arabidops
39	23	2.0	1830	8	AY063974	AY063974 Arabidops
40	23	2.0	1874	8	AY056387	AY056387 Arabidops
41	23	1.9	1363	4	PIGPEPA	M57267 Human proch
42	22	1.9	1398	4	PIGPEPA	J04601 Pig pepsino
43	22	1.9	1695	8	RMASPV	M20920 Swine peps1
44	22	1.9	1695	8	RMASPV	X56993 R.niveus ge
45	22	1.9	1695	8	RMASPV	X56993 R.niveus ge

ALIGNMENTS

RESULT 1
AX088019 LOCUS AX088019 1173 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 1 from Patent WO0114571.
ACCESSION AX088019
VERSION AX088019.1 GI:13396947

SOURCE

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE

AUTHORS van Rooijen, G., Keen, R. G., Boothe, J. and Shen, Y.
TITLE Commercial production of chymosin in plants
JOURNAL Patent: WO 0114571-A 1 01-MAR-2001;
Sembiosys Genetics Inc. (CA)

FEATURES

source 1. 1173
/organism="Bos taurus"
/db_xref="taxon:9913"
1. 1173
/note="unnamed protein product"

[illegible][illegible]

121 aagcgctgaagaacatggaattcttagaagattcttcagaacaacaglatgcatc 180
1674 AAGGGGCGTGAAGAGACATGAGACTTCTAGAGACTTCTTCACAACACAGATGGCACC 1733
181 agcagaagatctccggttcggtagaagttgcttagcgtgacattacccaacttctg 240
1734 AGCACAAGTACTCCGGCTTCGGTGAAGTTGCTAGCGTCCACTTACCAACTACCTTGAT 1793
241 agtaatactttgggaagatctactcggaaccccgctcaagagttacccgttctctt 300
1794 AGTCATATCTTTGGGAAGATCTACCTCGGAACCCCGCTCAAGAGTTCACCGTTCTT 1853
301 gatactgtctccctgaattctgggttccctctatctacttgcaagaacatgcttgaag 360
1854 GATACTGGTCTCTCGACTCTCGGGTTCCTCTATCTACTGCAAGACATGCTTCAG 1913
361 aaccaccaaagattcgatccgagaagaagtcgacccttcagaacttaggcaaaccttg 420
1914 AACCCACCAAGATTCGATCCGAGAAAGTCGTCACCTTCACAAGCTTAGGCAAAACCTTG 1973
421 tctatactacgtagacaggttagcagaagaatccttaggtatgataccgtctctg 480
1974 TCTATACACTACGCTACAGGTAGCATGCAAGGAATCTTAGGCTATGATACCGTCACTGTC 2033
481 tccaacattgtgagatctcaacagacagtagaacttagcaccacaaagacaggtgatg 540
2034 TCACAACATTTGSGAATTCACACAGACAGTACGACTTAGCACCACCAAGACAGTATGTC 2093
541 ttcacctatgcagaattcgatgagcatccttgatgcataccatcgctcgctcgaag 600
2094 TTCACCTATGCAAGATTCGATGCGATCCTTGATGAGCATACCCATCGCTCGTCAGAG 2153
601 tacttgataacctgtgttgacaacatgataacgaaccttagttagtcaaaactgttc 660
2154 TACTGATACCTGTGTGTGACACATGATGACCAACACACACTGATGCTAAGACTGTGTC 2213
661 tcggtttacatggaagaatgagcagaagagacatgctcagcttggagactatgatcca 720
2214 TCGGTTACATGAGACAGGAATGGCCAGAGACATGCTACGCTTGAGACTTATGATCCA 2273
721 tctactacaagaatctcttcaactcggttccagtcactgtgacagcagtaactgtgcaatc 780
2274 TCTCTACTACACAGGATCTCTTCACTGCGGTTCCAGTCACTGTCACACAGTACGCAATTC 2333
781 actgtagcagatgtaacatcagcaggtgtgtgtgtgtgcatgtagaagtgatgtaagct 840
2334 ACTGAGACAGTGTACCATGACGCGGTGTGTGTGTGATGTGAAGGTGATGTCAAGCT 2393
841 atcttgaataccgtagcgtlccaagcttggtcggaactagcagcagacatctcaaatcag 900
2394 ATCTTGATACCGGTACGATCCCAAGCTGTGCGAAGCTAGCAGAGACATTTCTCAACATTCAG 2453
901 caagctatgtgagcacaacagaacagtagagttgacatgattcgacaacctt 960
2454 CAAGCTATTTGGGCGCACACAGAACAGTACGCTGATTTGACATGATTTGCACAACTT 2513
961 agctacagcctacaggtgtgtctttagatcaacgcaagaatgtaccacatgaccccttc 1020
2514 AGCTACATGCTCTACAGTTGTCTTTAGATCAACGGCAAGATGTACCCATGACCCCTTC 2573
1021 gccctatccagcagaatcaagaaggtctgcaacagtggaatccagaagtgagaacatctc 1080
2574 GCCTATACACAGCAGATCAAGGTTTGCACACAGTGTTCACAGTGAAGAACATTC 2633
1081 cagaataatgactctgggagatgtgttcatctgtgagttactacagcgtcttggacaggcc 1140
2634 CAGAATAATGATCTTGGGAGATGTGTTCATTCGTGAGTACTACAGCGTCTTTGACAGGCC 2693
1141 aacaacctcgttgggctagctaaagaatctga 1173
2694 AACAACTCTGTTGGCTAGCTAAAGCAATTTGA 2726

RESULT 3
LOCUS BOVCHMOC 1269 bp mRNA linear MAM 28-FEB-1994
DEFINITION Bovine chymosin C (rennin) mRNA, complete cds.
ACCESSION J000004
VERSION J000004.1 GI:162861
KEYWORDS chymosin; chymosin C; protease; rennin.
SOURCE Bos taurus calf fourth stomach mucosa cDNA to mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1269)
Nishimori, K., Kawaguchi, Y., Hidaka, M., Uozumi, T. and Beppu, T.
TITLE Nucleotide sequence of calf prorennin cDNA cloned in Escherichia
JOURNAL J. Biochem. 91, 1085-1088 (1982)
MEDLINE 82189915
COMMENT chymosin is the major proteolytic enzyme in the fourth stomach of the unweaned calf. two chromatographically different forms, a and b, of the enzyme and its precursor are known and a third form, represented by this sequence, seems likely. the presence of a termination codon (bases 27-29) within the prepropeptide coding sequence implies that this sequence might be a pseudogene mrna. or there may be an error in the cdna. the amino acid sequence deduced from this nucleotide sequence differs at three sites from bovine chymosin a (see bovchymoa) and at four sites from bovine chymosin b (see bovchymob).
FEATURES
source Location/Qualifiers
1..1269
/organism="Bos taurus"
/db_xref="taxon:9913"
/tissue_type="fourth stomach mucosa"
/dev_stage="calf"
<1..1269
/product="chymoc mRNA"
402..1130
/note="precursor"
/codon_start=1
/product="chymosin C"
/protein_id="AAA30449.1"
/db_xref="GI:457097"
/translation="MGILGDTVTVSNIVDIQQTGSLGTPDGPVFTYAFEDGILGM
APSLASESLPVPDMNMRHLVADDFRSVYMDRNGOESMLFLGIDPSYTGSLHW
PVTQOQWQFVDSYTGISGVVACGGOALIDPCTSLVPSPSDILNQALATON
QYDERDIDCNLSHPTVPEINRITPLPSATYTGDDGFCCTGPOSENHSOKWILG
DVFLREYSVDFRANMLVGLAKTI"
BASE COUNT 299 a 381 c 321 g 268 t
ORIGIN
Query Match 4.8%; Score 56; DB 4; Length 1269;
Best Local Similarity 100.0%; Pred. No. 1.2e-18;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 760 gtcgacagatctggcaattcactgtgacagtgatccacataagcgggtgtgtgt 815
717 GTGCAGAGTACTGGCAATTCAGTGTGACAGTGTCAATCAGCGGTGTGTGT 772
RESULT 4
LOCUS E00079 1098 bp RNA linear PAT 29-SEP-1997
DEFINITION E00079.1 GI:2168383
ACCESSION E00079.1
VERSION E00079.1 JP 1983032896-A/1.
KEYWORDS JP 1983032896-A/1.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1098)
Beppu, T., Uozumi, T. and Nishimori, K.

TITLE CONJUGATED PLASMID AND MICROORGANISM CONTAINING THE SAME
JOURNAL Patent: JP 1983032896-A 1 25-FEB-1983;
BEPPI TERUHIKO
COMMENT OS calf
PN JP 1983032896-A/1
PD 25-FEB-1983
PF 24-AUG-1981 JP 1981131631
PI BEPPI TERUHIKO, UOZUMI TAKESHI, NISHIMORI KATSUHIKO PC
C07H21/04,C12N1/00,C12N15/00,(C12N1/00,C12R1:19); CC
strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue_type=mucous membrane of forth stomach; CC
*source: clone=PCR 100 1;
FH Key Location/Qualifiers
FH CDS 1..1098
FT Location/Qualifiers
1..1098
/product='calf prorennin'
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 252 a 327 c 300 g 219 t
ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1098;
Best Local Similarity 100.0%; Pred. No. 7.5e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tttctcggttacatgacaggaatgcccagagagacatgctcacgct 704
|||||
Db 581 tttctcggtttacatgacaggaatgcccagagagacatgctcacgct 629

RESULT 5
E02341 1098 bp RNA linear PAT 29-SEP-1997
LOCUS E02341
DEFINITION cDNA sequence coding for infant bovine prorennin.
ACCESSION E02341.1 GI:2170576
VERSION E02341.1
KEYWORDS JP 1990109984-A/1.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 1098)
BEPPI, T., UOZUMI, T. and NISHIMORI, K.
TITLE COMPLEX PLASMID AND MICROORGANISM CONTAINING THE SAME
JOURNAL Patent: JP 1990109984-A 1 23-APR-1990;
BEPPI TERUHIKO
COMMENT OS Bovine
PN JP 1990109984-A/1
PD 23-APR-1990
PF 01-DEC-1988 JP 1988302176
PI BEPPI TERUHIKO, UOZUMI TAKESHI, NISHIMORI KATSUHIKO PC
C12N15/59,C12N1/21;
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue_type=stomach mucosa;
FH Key Location/Qualifiers
FH CDS 1..1098
FT Location/Qualifiers
1..1098
/product='Infant bovine prorennin'
/organism='Bos taurus'
/db_xref='taxon:9913'
BASE COUNT 252 a 327 c 300 g 219 t

ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1098;
Best Local Similarity 100.0%; Pred. No. 7.5e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tttctcggttacatgacaggaatgcccagagagacatgctcacgct 704
|||||
Db 581 tttctcggtttacatgacaggaatgcccagagagacatgctcacgct 629

RESULT 6
AF177290 1101 bp mRNA linear MAM 24-JAN-2000
LOCUS AF177290
DEFINITION Bubalus arnee bubalis prochimysin mRNA, complete cds.
ACCESSION AF177290
VERSION AF177290.1 GI:6739579
KEYWORDS
SOURCE water buffalo.
ORGANISM Bubalus bubalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bubalus.
1 (bases 1 to 1101)
Batish, V. K., Mukhopadhyay, U. K., Mohanty, A. K., Grover, S. and
Kuipers, O. P.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1999) Molecular Biology Unit, Animal
Biotechnology Centre, National Dairy Research Institute, G. T. Road,
Karnal, Haryana 132001, India
Location/Qualifiers
1..1101
/organism='Bubalus bubalis'
/sub_species='bubalis'
/db_xref='taxon:89462'
1..1101
/codon_start=1
/product='prochimysin'
/protein_id='AAF27315.1'
/db_xref='GI:6739580'
/translation='MAETRTPLCGKSLRKALKEHGLLEDFLOKQOYGVSSKSGFG
EVASVPPLNFIDSOYRFGKIYGPPOEFTYVFDGSSDFWVPSLYCKSNCKNORFD
PKRSTFONLKGPLSIRIGTSGMOGILGDTATVSNYVDIOQTGISTOPPGDYFYA
EFEDGILGMAYPVVQOYWFYDVSITISGVVACEGGCAI11DTGTSKLVGSSDLN1Q
QAIGATONQYGEFDIDCNLSYMPYVSEIENKMGYPLTPSAVTSODGFCYSGPSEN
RSQOMILGDVFRREYVSFDRANLVGLAKAI'
BASE COUNT 244 a 336 c 306 g 215 t
ORIGIN

Query Match 4.2%; Score 49; DB 4; Length 1101;
Best Local Similarity 100.0%; Pred. No. 7.5e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tttctcggttacatgacaggaatgcccagagagacatgctcacgct 704
|||||
Db 584 tttctcggtttacatgacaggaatgcccagagagacatgctcacgct 632

RESULT 7
SYNPROCA 1115 bp DNA linear SYN 27-APR-1993
LOCUS SYNPROCA
DEFINITION Synthetic bovine prochimysin A protein region.
ACCESSION M22593 M18758
VERSION M22593.1 GI:209139
KEYWORDS prochimysin.
SOURCE Synthetic DNA.
ORGANISM Synthetic construct
REFERENCE 1 (bases 1 to 1115)
Wosnick, M. A., Barnett, R. W., Vicentini, A. M., Erfle, H., Elliott, R.,

TITLE Summer-Smith,M., Mantel,N. and Davies,R.W.
Rapid construction of large synthetic genes: total chemical synthesis of two different versions of the bovine prochymosin gene

JOURNAL Gene 60, 115-127 (1987)

MEDLINE 88152494

FEATURES
source Location/Qualifiers
1. 1115 /organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 254 a 332 c 302 g 227 t

ORIGIN 1 bp downstream of HindIII site.

Query Match 4.2%: Score 49; DB 12; Length 1115;
Best Local Similarity 100.0%: Pred. No. 7.5e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 656 tttctcggttacatgacaggaatggccagagagacatgctcacgct 704
|||||
Db 580 TGTTCGCTTACATGACAGATGGCCAGAGAGCATGCTCACGCT 628

RESULT 8
LOCUS 104058 1175 bp linear PAT 02-DEC-1994
DEFINITION Sequence 5 from Patent EP 0123928.
ACCESSION 104058
VERSION 104058.1 GI:591912
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1175)
AUTHORS Gashon,L.M., McCaman,M.T., Rice,C.W. and Sias,S.R.
TITLE Recombinant DNA coding for a polypeptide displaying milk clotting activity
JOURNAL Patent: EP 0123928-A2 5 07-NOV-1984;
FEATURES
source Location/Qualifiers
1. 1175 /organism="unknown"

BASE COUNT 263 a 356 c 318 g 238 t

ORIGIN

Query Match 4.2%: Score 49; DB 6; Length 1175;
Best Local Similarity 100.0%: Pred. No. 7.4e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 656 tttctcggttacatgacaggaatggccagagagacatgctcacgct 704
|||||
Db 649 TGTTCGCTTACATGACAGATGGCCAGAGAGCATGCTCACGCT 697

RESULT 9
LOCUS AR002347 1240 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 2 from patent US 5741665.
ACCESSION AR002347
VERSION AR002347.1 GI:3963901
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1240)
AUTHORS Kato,E.K. and Stuart,W.Dorsey.
TITLE Light-regulated promoters for production of heterologous proteins in filamentous fungi
JOURNAL Patent: US 5741665-A 2 21-APR-1998;
FEATURES
source Location/Qualifiers
1. 1240 /organism="unknown"

BASE COUNT 274 a 374 c 339 g 253 t

ORIGIN

Query Match 4.2%: Score 49; DB 6; Length 1240;
Best Local Similarity 100.0%: Pred. No. 7.4e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 656 tttctcggttacatgacaggaatggccagagagacatgctcacgct 704
|||||
Db 700 TGTTCGCTTACATGACAGATGGCCAGAGAGCATGCTCACGCT 748

RESULT 10
LOCUS BOVCHYMOA 1275 bp mRNA linear MAM 26-APR-1993
DEFINITION bovine chymosin a (rennin) mRNA.
ACCESSION J00002
VERSION J00002.1 GI:162857
KEYWORDS chymosin; chymosin A; rennin.
SOURCE bovine (calf) cdna of fourth stomach mucosa mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1275)
AUTHORS Molr,D., Mao,J.L., Schumm,J.W., Vovis,G.F., Alford,B.L. and Taunton-Rigby,A.
TITLE molecular cloning and characterization of double-stranded cdna coding for bovine chymosin
JOURNAL Gene 19, 127-138 (1982)
MEDLINE 83054629
COMMENT chymosin (rennin) is the major proteolytic enzyme in the fourth stomach of the unweaned calf. two chromatographically distinct forms are known and a third seems likely (see bovchymob and bovchymoc). this sequence has been tentatively identified as chymosin a; it differs from chymosin b by only two amino acids and from chymosin c by another amino acid. [1] argues that the different chymosins are probably polymorphic variants of a single chymosin gene.

FEATURES
source Location/Qualifiers
1. 1275 /organism="Bos taurus"
/db_xref="taxon:9913"
21. 1166 /note="preprochymosin a"
/protein_id="AAA30447.1"
/db_xref="GI:162858"
/translation="MRCILVYLAVFALSGAETRIPIYKKSIRKALKHGLEDFLQKOYGISISKYSGFEVASVPLTNVLDISOYFGKLYIGTPPOEFVLPDGTSSDWPVS IYCKSNCKNQRQFDPKRSFQNLGRLPLSTHYGTSGMOGLIGDYTVASIIIDIOOT VGLSTQEPGVDFYAFEDGILGMAVPSLASFSYIPVDNMNRLVADOLFESYMDRN GOESMLTGLAIDPEYITGSLHWVYTVQOIKQFTVDSVTTISGVYVACEGCOALIDTG TSKLVGPSSDILNIOAIGATONQYDFEDIDCNLSYMPVFEINCKMPLTPSAVT SODGFCFTSGFQSENHSOKWILGDVFTREYYSVDRANNNLGLAKAI"

sig_peptide 24. 68
mat_peptide 195. 1163
/product="chymosin a"

BASE COUNT 293 a 391 c 336 g 255 t

ORIGIN 20 bases upstream from codon 1.

Query Match 4.2%: Score 49; DB 4; Length 1275;
Best Local Similarity 100.0%: Pred. No. 7.3e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 656 tttctcggttacatgacaggaatggccagagagacatgctcacgct 704
|||||
Db 649 TGTTCGCTTACATGACAGATGGCCAGAGAGCATGCTCACGCT 697

RESULT 11
E00144

LOCUS E00144 1289 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding bovine calf chymosin.
ACCESSION E00144
VERSION E00144.1 GI:2168443
KEYWORDS JP 1984021392-A/1.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1289)
AUTHORS Chibaarsu, E.B.
TITLE CATTLE KIMOCIN
JOURNAL Patent: JP 1984021392-A 1 03-FEB-1984;
GENEX CORP
COMMENT OS bovine calf
PN JP 1984021392-A/1
PD 03-FEB-1984
PE 30-JUN-1983 JP 1983119481
PR 01-JUL-1982 US 82 394433, 13-APR-1983 US 83 484539 PI
CHIVARUSU EI BASURETSUTO
PC C12N15/00, C07H21/04, C12M1/20, C12P19/34, C12P21/02, (C12N15/00,
C12R1:19);
CC strandedness: Double;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue, type=stomach mucosa;
FH Key Location/Qualifiers
FT mat_peptide 169..1113
FT /product='bovine calf chymosin' FT CDS
FT <1..1116 /product='precursor polypeptide'.
FEATURES
source Location/Qualifiers
1..1289
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 311 a 388 c 329 g 260 t 1 others
ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1289;
Best Local Similarity 100.0%; Pred. No. 7.3e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tgtctcggttacatgacaggaatgcccagagagatgctcagct 704
|||||
Db 623 tgttctcggttacatgacaggaatgcccagagagatgctcagct 671

RESULT 12
A15836 A15836 1291 bp DNA linear PAT 16-FEB-1994
LOCUS chymosin gene.
DEFINITION A15836
ACCESSION A15836
VERSION A15836.1 GI:488959
KEYWORDS
SOURCE .
ORGANISM cow.
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1291)
AUTHORS Simons, A.F.M. and De Vos, W.M.
TITLE DNA fragments, containing a lactic acid bacterium-specific
regulator region for the expression of genes coding for normally
heterologous proteins
JOURNAL Patent: EP 0307011-A 5 15-MAR-1989;
NEDERLANDS INSTITUUT VOOR ZUIVELONDERZOEK
FEATURES
source Location/Qualifiers
1..1291
/organism='Bos taurus'
/db_xref='taxon:9913'
BASE COUNT 323 a 382 c 328 g 257 t 1 others

ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1291;
Best Local Similarity 100.0%; Pred. No. 7.3e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tgtctcggttacatgacaggaatgcccagagagatgctcagct 704
|||||
Db 609 tgttctcggttacatgacaggaatgcccagagagatgctcagct 657

RESULT 13
BOVCHYMOB
LOCUS BOVCHYMOB 1305 bp mRNA linear MAM 26-APR-1993
DEFINITION bovine chymosin b (rennin) mRNA.
ACCESSION J00003
VERSION J00003.1 GI:162859
KEYWORDS chymosin; chymosin B; protease; rennin.
SOURCE bovine (calf) cDNA of fourth stomach mucosa mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1305)
AUTHORS Harris, T.J.R., Lowe, P.A., Lyons, A., Thomas, P.G., Eaton, M.A.W.,
Mallican, T.A., Patel, T.P., Bose, C.C., Carey, N.H. and Doel, M.T.
TITLE molecular cloning and nucleotide sequence of cDNA coding for calf
preprochymosin
JOURNAL Nucleic Acids Res. 10, 2177-2187 (1982)
MEDLINE 82221400
COMMENT chymosin is the major proteolytic enzyme in the fourth stomach of
the unweaned calf. Two chromatographically different forms, a and
b, of the enzyme and its precursor are known and a third form seems
likely (see bovcymoa, bovcymoc). This sequence has tentatively
been identified as coding for preprochymosin b. Sequence comparison
indicates that the precursors for chymosins a and b differ by only
two amino acids, and for b and c by only four amino acids.

FEATURES
source Location/Qualifiers
1..1305
/organism='Bos taurus'
/db_xref='taxon:9913'
<1..1305
/product='chymob mRNA'
26..1171
/note='preprochymosin b'
/codon.start=1
/protein.id='AAA30448.1'
/db_xref='GI:162860'
/translation='MRLVYLAVFALSQGAELTRIPYKSKSRALKRHLGLEDPL
QKQDYGISKRSYSGFEVAVPLNTYLDQSFGKLYLSTPPOEFTVLEDTGSSDPWVS
IYCKSNACKNHQRFDPKRSSTFQMLGRPLIHVGTGSMQILGDTVTSNIVDIQT
VGLSTQEPGDVFTYAEFDGILGMAYPSLAESYSPVDMMNRHLVADLFVYMDRN
GOESMLTGLAINPSYTGSLHWVGVQVOXQVSTVDSVTLSGVVAVACGGCALIDTG
TSKRGSSDILINPSTIGALNIGATONXGFEFDIDCNLSYMPVPEINKNMPLPSAAT
SDDGFGCTSGFQSEHSQKMLGDVFIREFYVFRANNLVGLAKAI'
29..73
/note='presequence'
mat_peptide 200..1168
/product='chymosin b'
BASE COUNT 305 a 393 c 340 g 267 t
ORIGIN pst-1 site.

Query Match 4.2%; Score 49; DB 4; Length 1305;
Best Local Similarity 100.0%; Pred. No. 7.3e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tgtctcggttacatgacaggaatgcccagagagatgctcagct 704
|||||
Db 654 tgttctcggttacatgacaggaatgcccagagagatgctcagct 702

RESULT 14
E00042 1460 bp RNA linear PAT 29-SEP-1997
LOCUS E00042
DEFINITION DNA coding of pre-pro rennin.
ACCESSION E00042
VERSION E00042.1 GI:2168348
KEYWORDS JP 1982141287-A/1.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1460)
AUTHORS Baanadetsute,R.A., Jient,M., Donarudo,T.M., Arison,T.R. and Jierarudo,F.B.
TITLE RENININ, PREPRORENININ OR PRORENININ GENE OBTAINED FROM RECOMBINED DNA MATERIAL AND LIVE CELL CONTAINING GENE
JOURNAL Patent: JP 1982141287-A 1 01-SEP-1982;
COMMENT KORABORATEIBU RES INC
OS call
PN JP 1982141287-A/1
PD 01-SEP-1982
PF 14-JAN-1982 JP 1982003556
PR 16-JAN-1981 US 81 225717, 01-DEC-1981 US 81 325481 PI
BAANADETSUTE RABUTSUTSUKII AKUFUODO, JIENT MAO, PI DONARUDO TEIRAA MOIAA,
PI ARISON TAUNTON RIGUBII, JIERARUDO FURANSHISU BUBUISU PC
C12N1/00,C07G7/00,C07H21/04,C12N15/00,C12P21/00//C12N1/18, PC
C12N1/20,
PC C12P19/34,C12R1/19,C12R1/865;
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue_type=stomach;
FH key Location/Qualifiers
FT CDS 205..1350
FT /product='pre-pro rennin'
FT mat_peptide 205..1347
FT /product='pre-pro rennin'
FT CDS 253..1350
FT /product='pro rennin'
FT mat_peptide 253..1347
FT /product='pro rennin'
FT CDS 278..1350
FT /product='pro rennin'
FT mat_peptide 278..1347
FT /product='rennin'
FEATURES
source Location/Qualifiers
1..1460
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 328 a 440 c 398 g 294 t
ORIGIN
Query Match 4.2%; Score 49; DB 6; Length 1460;
Best Local Similarity 100.0%; Pred. No. 7.2e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 656 tggttcggtttacatgacaggaatgccaagagagcatgctcacgct 704
|||||
Db 833 tggttcggtttacatgacaggaatgccaagagagcatgctcacgct 881
RESULT 15
E00295 1460 bp RNA linear PAT 29-SEP-1997
LOCUS E00295
DEFINITION cDNA encoding pre-prorennin A.
ACCESSION E00295
VERSION E00295.1 GI:2168583
KEYWORDS JP 1985058077-A/3.
SOURCE cow.
ORGANISM Bos taurus

REFERENCE 1 (bases 1 to 1460)
AUTHORS Baanadetsute,R.A., Jient,M., Donarudo,T.M., Arison,T.R., Robaato,J.N., Jien,I.M., Donarudo,T.M. and Kurisutofuna,G.G.
TITLE USE OF GAL YEAST PROMOTOR
JOURNAL Patent: JP 1985058077-A 3 04-APR-1985;
COMMENT KORABORATEIBU RES INC
OS Bovine
PN JP 1985058077-A/3
PD 04-APR-1985
PF 28-FEB-1984 JP 1984035472
PR 28-FEB-1983 US 83 470911
PI DEBUTSUDO BOTSUTOSUTREIN, RONARUDO UETIN DEIBUISU, PI JIERARUDO RARUFU FUIINKU,
PI ARISON TAUNTON RIGUBII, ROBAATO JIENTORII NOURUTON, JIENT I MAO, PI DONARUDO TEIRAA MOA, KURISUTOFUNA GONSUDOFURII GOFU PC
C12N15/00,C07H21/04,C07K13/00,C12N1/16,C12P21/02,C12N1/16, PC
C12R1/865),
PC (C12P21/02,C12R1/865);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue_type=stomach mucosa;
FH key Location/Qualifiers
FT CDS 205..1350
FT /product='pre-prorennin A'
FT mat_peptide 205..1347
FT /product='pre-prorennin A'
FT mat_peptide 253..1347
FT /product='Prorennin'
FEATURES
source Location/Qualifiers
1..1460
/organism='Bos taurus'
/db_xref='taxon:9913'
BASE COUNT 327 a 437 c 400 g 296 t
ORIGIN
Query Match 4.2%; Score 49; DB 6; Length 1460;
Best Local Similarity 100.0%; Pred. No. 7.2e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 656 tggttcggtttacatgacaggaatgccaagagagcatgctcacgct 704
|||||
Db 833 tggttcggtttacatgacaggaatgccaagagagcatgctcacgct 881

Search completed: August 1, 2002, 01:41:47
Job time: 16797 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 21:15:05 : Search time 791.3 Seconds
(without alignments)
2545.104 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173
Sequence: 1 atggaacttccttaagtcctt.....ggctagctaagaacatctga 1173

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size: 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N_Geneseq_032802.*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1173	100.0	1173	22	AAS00569
2	1173	100.0	3957	22	AAS00570
3	49	4.2	1098	4	AA030063
4	49	4.2	1098	11	AA040683
5	49	4.2	1175	5	AA040295
6	49	4.2	1175	13	AA020949
7	49	4.2	1210	12	AA014051
8	49	4.2	1240	16	AA030065
9	49	4.2	1278	5	AA040055

10	49	4.2	1289	4	AA030022
11	49	4.2	1291	10	AA091157
12	49	4.2	1314	4	AA030049
13	49	4.2	1460	3	AA020043
14	49	4.2	1460	5	AA040180
15	49	4.2	1460	10	AA040188
16	49	4.2	2733	20	AA020643
17	49	4.2	2982	10	AA091185
18	38	3.2	1290	4	AA030209
19	35	3.0	637	14	AA040459
20	34	2.9	1096	20	AA083966
21	33	2.8	179	5	AA040296
22	33	2.8	186	13	AA020950
23	33	2.8	819	20	AA083967
24	32	2.7	107	12	AA014777
25	29	2.5	1460	3	AA020043
26	29	2.5	1460	5	AA040180
27	23	2.0	1083	21	AA044658
28	23	2.0	1140	24	AA097159
29	23	2.0	1814	21	AA037875
30	20	1.7	828	11	AA003224
31	20	1.7	846	11	AA003223
32	20	1.7	1104	9	AA080001
33	20	1.7	1143	5	AA040214
34	20	1.7	11624	19	AA052850
35	19	1.6	231	23	AA057374
36	19	1.6	689	19	AA059757
37	19	1.6	1353	21	AA039862
38	19	1.6	1751	19	AA059610
39	19	1.6	1774	20	AA042126
40	19	1.6	3362	18	AA06278
41	19	1.6	3747	18	AA06250
42	19	1.6	4049	22	AA02878
43	19	1.6	4170	22	AA009546
44	18	1.5	158	21	AA021764
45	18	1.5	430	22	AA060501

ALIGNMENTS

RESULT 1	
AAS00569	standard; DNA: 1173 Bp.
AC	AAS00569;
XX	
DT	14-MAY-2001 (first entry)
XX	
DE	Bovine pre-pro-chymosin DNA sequence.
XX	
KW	Chymosin; transcription regulator; terminator sequence; soybean; corn;
KW	pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;
KW	barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;
KW	safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;
KW	squash; jojoba; ds.
XX	
OS	Bos sp.
XX	
FH	Key
FT	1..1173
FT	CDS
FT	
FT	sig_peptide
FT	1..78
FT	/tag= a
FT	/product= "Bovine chymosin"
FT	
FT	misc_feature
FT	79..201
FT	/tag= c
FT	/note= "Pro sequence"
FT	202..1170
FT	/tag= d
FT	/product= "Mature bovine chymosin"
XX	
PN	MO200114571-A1.

Sequence of veal c
Cloned sequence of
CDNA sequence corr
Pre-prorenin-A ge
Sequence of recomb
BamHI/SalI insert
2.7 Kbp HindIII fr
BamHI insert from
Sequence of prepro
Prochymosin gene 5
DNA encoding the f
Sequence coding fo
DNA coding for a r
DNA encoding His-c
Renin - casein co
Pre-prorenin-A ge
Sequence of recomb
Arabidopsis thalia
Human aspartyl pro
Arabidopsis thalia
DNA fragment of pa
Optimised prochymo
Gene encoding (pre
Human eyal gene co
cDNA #50 encoding
Human secreted pro
Plasmodium ovale p
Human secreted pro
Human normal blad
Human putrocin-se
Human shear stress
Human protease pro
Human secreted pro
Human immune/haema

XX 01-MAR-2001.
 PD 23-AUG-2000; 2000MO-CA00975.
 PF 23-AUG-1999; 99US-0378696.
 PR (SEMB-) SEMBIOSYS GENETICS INC.
 PA Van Rooijen G, Keon RG, Boothe J, Shen Y;
 PI WPI; 2001-226621/23.
 DR P-PSDB; AA000536.
 XX Producing chymosin in seeds of plants such as rice, flax, rape seed, by
 PT transforming plant cell with a nucleic acid encoding chymosin operably
 PT linked to transcription regulator and terminator sequences -
 XX Claim 9; Fig 1; 56pp; English.
 PS The sequence represents a DNA which encodes a bovine chymosin
 CC polypeptide. Chymosin can be produced in a plant seed through
 CC introduction of a chimeric nucleic acid molecule, comprising a nucleic
 CC acid sequence encoding a chymosin polypeptide operatively linked to
 CC transcription regulator and terminator sequences, into a plant cell. The
 CC sequences are useful for producing plant seeds, in particular seeds of
 CC soybean, rape seed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
 CC barley, oats, sorghum, Arabidopsis thaliana, potato, flax/linseed,
 CC safflower, oil palm, groundnut, Brazil nut, coconut, castor, coriander,
 CC squash, jojoba and rice.
 CC
 XX Sequence 1173 BP; 299 A; 308 C; 262 G; 304 T; 0 other;
 SQ

Query Match 100.0%; Score 1173; DB 22; Length 1173;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaattccttaagcttccttccttctactagcttccttccttcgttcgtcaataactcgtt 60
 DB 1 atgaattccttaagcttccttccttctactagcttccttccttcgttcgtcaataactcgtt 60
 QY 61 gctgttactaacgctcgtcgtgagatcacccgcatctcctctctaaaggtaagcttcgtt 120
 DB 61 gctgttactaacgctcgtcgtgagatcacccgcatctcctctctaaaggtaagcttcgtt 120
 QY 121 aagcgctgaagaagcagcttctagaagaactctcttcgacagaacaacagatgcatc 180
 DB 121 aagcgctgaagaagcagcttctagaagaactctcttcgacagaacaacagatgcatc 180
 QY 181 agcagaagctacccgcttcggtgaagtttgtagcgttgccacttaccactacttgat 240
 DB 181 agcagaagctacccgcttcggtgaagtttgtagcgttgccacttaccactacttgat 240
 QY 241 agtcaatacttgggaagaatctactctcggaaccccgctccaagaagttaccgtttcttt 300
 DB 241 agtcaatacttgggaagaatctactctcggaaccccgctccaagaagttaccgtttcttt 300
 QY 301 gatctgtgttccttcgacttcgttggtttccctctactacttgtaagaagcaatcctgcaag 360
 DB 301 gatctgtgttccttcgacttcgttggtttccctctactacttgtaagaagcaatcctgcaag 360
 QY 361 aaccaccaaagaattcgatccgagaagtcgtccacacttccaaactttaggcaaaccttg 420
 DB 361 aaccaccaaagaattcgatccgagaagtcgtccacacttccaaactttaggcaaaccttg 420
 QY 421 tctataactacggtacaggtagcatgacgaagaatcttaggctatgataccgtcactgttc 480
 DB 421 tctataactacggtacaggtagcatgacgaagaatcttaggctatgataccgtcactgttc 480
 QY 481 tccacaattgtgacattcaacagacagataggaacttagaccacaaacaggttagtc 540
 DB 481 tccacaattgtgacattcaacagacagataggaacttagaccacaaacaggttagtc 540

QY 541 ttcacctaagcagaatctcgatgcatccttcgtgtatggcatatcccatcgctcgtgcagag 600
 DB 541 ttcacctaagcagaatctcgatgcatccttcgtgtatggcatatcccatcgctcgtgcagag 600
 QY 601 tactcgaatacctgtgtttgacaacatgatgaaccgacacctaagtagctcaagactttc 660
 DB 601 tactcgaatacctgtgtttgacaacatgatgaaccgacacctaagtagctcaagactttc 660
 QY 661 tcggtttacatgagcagaatgagcagaagagacatgctcaacgcttggagctatgtacca 720
 DB 661 tcggtttacatgagcagaatgagcagaagagacatgctcaacgcttggagctatgtacca 720
 QY 721 tccctacacaaagatctcttcacatggttcccaactcgtgacagcagtagctgcaattc 780
 DB 721 tccctacacaaagatctcttcacatggttcccaactcgtgacagcagtagctgcaattc 780
 QY 781 actgtgacaggtgcacatcagcgtgtgtgtttgttcgcatgtgaaggtgagatcgaagct 840
 DB 781 actgtgacaggtgcacatcagcgtgtgtgtttgttcgcatgtgaaggtgagatcgaagct 840
 QY 841 atcttgataccggttacgtlccaaagctgtgcgacctagcagcagcatctcaacatcag 900
 DB 841 atcttgataccggttacgtlccaaagctgtgcgacctagcagcagcatctcaacatcag 900
 QY 901 caagctatgtgagccacacagaacacagatgaggtgtgttgatagatgtgacaaactt 960
 DB 901 caagctatgtgagccacacagaacacagatgaggtgtgttgatagatgtgacaaactt 960
 QY 961 agctacatgctacagttgtctttgagatacagcagcagaagatgataccacttacccttc 1020
 DB 961 agctacatgctacagttgtctttgagatacagcagcagaagatgataccacttacccttc 1020
 QY 1021 gctctataccagcagatcaagaggttctgcacacagttgattccagatgtgagaacatcc 1080
 DB 1021 gctctataccagcagatcaagaggttctgcacacagttgattccagatgtgagaacatcc 1080
 QY 1081 cagaatgagcttggggagatgtgttcatttggtagtactacaaggttcttgacagggcc 1140
 DB 1081 cagaatgagcttggggagatgtgttcatttggtagtactacaaggttcttgacagggcc 1140
 QY 1141 aacaacctcgttggcgtagctaaagcaatctga 1173
 DB 1141 aacaacctcgttggcgtagctaaagcaatctga 1173

RESULT 2
 AAS00570
 ID AAS00570 standard; DNA: 3957 BP.
 XX
 AC AAS00570;
 XX
 DT 14-MAY-2001 (first entry)
 XX
 DE Bovine phaseolin promoter pre-pro-chymosin-phaseolin terminator.
 XX
 KW Chymosin; transcription regulator; terminator sequence; soybean; corn;
 KW pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;
 KW barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;
 KW safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;
 KW squash; jojoba; ds; phaseolin; promoter; terminator; mutant; French bean.
 XX
 OS Chimeric - Bos sp.
 OS Chimeric - Phaseolus vulgaris.
 OS
 XX
 FH Key
 FT Location/Qualifiers
 FT 1..1553
 FT /*tag= a
 FT /note= "Phaseolin promoter"
 FT CDS 154..2726
 FT /*tag= b
 FT /product= "Bovine pre-pro-chymosin"
 FT 2727..3957
 FT terminator

```
FT      /*tag= C
FT      /note= "Phaseolin terminator"
XX      MO200114571-A1.
XX      01-MAR-2001.
XX      23-AUG-2000; 2000MO-CA00975.
XX      23-AUG-1999; 99US-0378696.
XX      (SEMB-) SEMBIOSIS GENETICS INC.
XX      Van Rooijen G, Keon RG, Boothe J, Shen Y;
XX      WPI: 2001-226621/23.
XX      P-PSDB: AAU00536.
XX      Producing chymosin in seeds of plants such as rice, flax, rape seed, by
PT      transforming plant cell with a nucleic acid encoding chymosin operably
PT      linked to transcription regulator and terminator sequences -
XX
PS      Example 1; Fig 2; 56pp; English.
CC      The sequence represents a chimeric polynucleotide comprising a
CC      pre-pro-chymosin, a phaseolin promoter and a phaseolin terminator.
CC      Chymosin can be produced in a plant seed through introduction of a
CC      chimeric nucleic acid molecule, comprising a nucleic acid sequence
CC      encoding a chymosin polypeptide operatively linked to transcription
CC      regulator and terminator sequences, into a plant cell. The sequences are
CC      useful for producing plant seeds, in particular seeds of soybean, rape
CC      seed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats,
CC      sorghum, Arabidopsis thaliana, potato, flax/linseed, safflower, oil palm,
CC      groundnut, Brazil nut, coconut, castor, coriander, squash, jojoba and
CC      rice.
SQ      Sequence 3957 BP; 1263 A; 790 C; 609 G; 1295 T; 0 other:
Query Match      100.0%; Score 1173; DB 22; Length 3957;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 atgaattccttaagctcttcctccttctaagcttccttccttggttcgtaatacttcgtt 60
DB      1554 atgaattccttaagctcttcctccttctaagcttccttccttggttcgtaatacttcgtt 1613
QY      61 gctgttaactcaagctgtgtgagatacccgcatctcctcttacaagaagtaagtcctcgt 120
DB      1614 gctgttaactcaagctgtgtgagatacccgcatctcctcttacaagaagtaagtcctcgt 1673
QY      121 aaggcgctgaaggaacatgagactttagaagactctctgtgcaaaacacagatgcatc 180
DB      1674 aaggcgctgaaggaacatgagactttagaagactctctgtgcaaaacacagatgcatc 1733
QY      181 agcagaagtaactccggtctcggtgaagtgttgtagcggtgcaccttaccacttcgtat 240
DB      1724 agcagaagtaactccggtctcggtgaagtgttgtagcggtgcaccttaccacttcgtat 1793
QY      241 agtcaatacttgggaagatctaactcggaaacccgcgcctcaagaagttcacggtctctt 300
DB      1794 agtcaatacttgggaagatctaactcggaaacccgcgcctcaagaagttcacggtctctt 1853
QY      301 gatactggttctctgacttgggttgggttcctctatctatctatgnaaagaatgctgcgaag 360
DB      1854 gatactggttctctgacttgggttgggttcctctatctatctatgnaaagaatgctgcgaag 1913
QY      361 aaccaccaaagatcgcagcagagaagtcgctcacacttccacaactttaggcaaaccttg 420
DB      1914 aaccaccaaagatcgcagcagagaagtcgctcacacttccacaactttaggcaaaccttg 1973
QY      421 tctataactacggtacaggttagcatgcaaggaatcttaagctatgatacgttcactg 480
DB      tctataactacggtacaggttagcatgcaaggaatcttaagctatgatacgttcactg 480
```

```
DB      1974 tctataactacggtacaggttagcatgcaaggaatcttaagctatgatacgttcactg 2033
QY      481 tccaattgttggaatttcaacagacagtagacttagagaccacgaagatgtagtc 540
DB      2034 tccaattgttggaatttcaacagacagtagacttagagaccacgaagatgtagtc 2093
QY      541 tccaactacgaagaatcgcagtccttcgtgatgtagcaaccacgtctgcgtcaag 600
DB      2094 tccaactacgaagaatcgcagtccttcgtgatgtagcaaccacgtctgcgtcaag 2153
QY      601 taactgatacctgtgttgaacaacatgatacgaacccagacttagctcaagactgttc 660
DB      2154 taactgatacctgtgttgaacaacatgatacgaacccagacttagctcaagactgttc 2213
QY      661 tcggtttacatggaaggaatggccagagagacatgcttaagcttggaaattattatcca 720
DB      2214 tcggtttacatggaaggaatggccagagagacatgcttaagcttggaaattattatcca 2273
QY      721 tctactacacaggaatcttcaactgtggttcagatcactgttcaagcagtagctgcaatc 780
DB      2274 tctactacacaggaatcttcaactgtggttcagatcactgttcaagcagtagctgcaatc 2333
QY      781 actgttgacagtgatcaacatcagcggtgtgtgtgtgcatgtgaagtgatgtcaagct 840
DB      2334 actgttgacagtgatcaacatcagcggtgtgtgtgtgcatgtgaagtgatgtcaagct 2393
QY      841 atcttggataccggtgaagccaaagctgtgtcggaaccttagcagagacatttcaacattcg 900
DB      2394 atcttggataccggtgaagccaaagctgtgtcggaaccttagcagagacatttcaacattcg 2453
QY      901 caagcatatggagccacacagaacacagtagcaggtgaagttgacatagatgcgacaacct 960
DB      2454 caagcatatggagccacacagaacacagtagcaggtgaagttgacatagatgcgacaacct 2513
QY      961 agctatactcctacagtgcttctttagatcaacagcgaagatgtaccactgtgaccccttc 1020
DB      2514 agctatactcctacagtgcttctttagatcaacagcgaagatgtaccactgtgaccccttc 2573
QY      1021 gctctacacagcgaagatgaaggttcttcaccagtggaattccagatgtagaacaatttc 1080
DB      2574 gctctacacagcgaagatgaaggttcttcaccagtggaattccagatgtagaacaatttc 2633
QY      1081 cagaatgagatcttggagatgtgttcattcgttgagtaactacagcgtctttagcagggcc 1140
DB      2634 cagaatgagatcttggagatgtgttcattcgttgagtaactacagcgtctttagcagggcc 2693
QY      1141 aacaacctcgttgggtagctaaagcaatctga 1173
DB      2694 aacaacctcgttgggtagctaaagcaatctga 2726
RESULT      3
AAN30063
ID      AAN30063 standard; cDNA; 1098 BP.
XX
AC      AAN30063;
XX
DE      14-JUN-1992 (first entry)
XX
DE      Sequence of prorennin cDNA in PCR 10001.
XX
KM      Rennin; renin; enzyme; protease; ss.
XX
OS      Bos taurus.
XX
FH      Key
FT      CDS
XX      Location/Qualifiers
XX      EP73029-A.
XX      1..1098
XX      /*tag= a
XX      02-MAR-1983.
XX
```



```

DE Prochymosin (prorennin) gene from calf.
XX
XX Rennet; zymogen; autocatalytic activation; pseudorennin; milk;
XX clotting activity; ss.
XX
XX Bos taurus.
OS
XX
XX Key Location/Qualifiers
XX CDS 21..1158
XX /*tag= a
XX /product= prorennin
XX /note= "also known as prochymosin"
XX
XX US5082775-A.
XX
XX 21-JAN-1992.
XX
XX 28-OCT-1988; 88US-0263927.
XX
XX 28-OCT-1988; 88US-0263927.
XX
XX 11-MAY-1984; 84US-0609495.
XX
XX 12-DEC-1986; 86US-0940199.
XX
XX 31-MAR-1983; 83US-0480860.
XX
XX 28-APR-1986; 86US-0856700.
XX
XX (BERL-) BERLEX LABS INC.
XX
XX McCaman MT, King JF;
XX
XX WPI; 1992-049149/06.
XX
XX P-PSDB; AAR20730.
XX
XX
XX Isolating heterologous polypeptide from bacterial inclusion
XX bodies - by lysing cells, extrn. with nonionic detergent and sepg.
XX
XX Insoluble polypeptide
XX
XX Disclosure; Fig 6; 21pp; English.
XX
XX The prorennin (prochymosin) sequence was obt'd. from mRNA from
XX unweaned calf's stomach (abomasum). To identify rennin-specific
XX mRNA species, hybridisation with rennin specific probes was performed
XX using probes derived from the published amino acid sequence of
XX prorennin (Follmann et al., J. Biol. Chem. 254, 8447-8456 (1979))
XX (see AAQ3321, 2). Reverse transcriptase was used to transcribe mRNA
XX into a cDNA copy which was inserted into plasmid pBR322 and cloned
XX into E. coli strain K-12 MM 294 (ATCC 31446). The prorennin
XX specific clones were used to screen for recombinant plasmids.
XX Only two clones 563 and 15C5 were found to contain the whole prorennin
XX sequence, with clone 565 having its complete coding sequence (shown
XX here). The sequence was the same as the published sequence except
XX for two silent mutations in codons 274 and 336 respectively, and a
XX single mutation in codon 302 which converts the sequence to the B form
XX of rennin. The product of prorennin, rennin is an active component of
XX rennet which is used to clot milk in the process of making cheese.
XX See also AAQ20950.
XX
XX
XX Sequence 1175 BP; 261 A; 355 C; 319 G; 240 U; 0 other;
XX
XX
XX Query Match 4.2%; Score 49; DB 13; Length 1175;
XX Best Local Similarity 75.5%; Pred. No. 1e-14;
XX Matches 37; Conservative 12; Mismatches 0; Indels 0; Gaps 0;
XX
XX 656 tgtctcggttacatgacaggaatgcccagagagacatgctcagct 704
XX :||||:||||:||||:||||:||||:||||:||||:||||:||||:
XX Db 649 ugucucgguuacagagacaggaugccagagagacagcucagcu 697
XX
XX
XX RESULT 7
XX ID AAQ14051 standard; DNA: 1210 BP.
XX
XX AC AAQ14051;

```

```

XX
XX 06-JAN-1992 (first entry)
XX
XX Rennin gene.
XX
XX
XX Prorennin; alpha-S1-casein gene; insulin-like growth factor I; IGF-I;
XX mammary gland; ss.
XX
XX Synthetic.
XX
XX EP451823-A.
XX
XX 16-OCT-1991.
XX
XX 10-APR-1991; 91EP-0105702.
XX
XX 19-APR-1990; 90DE-4012526.
XX
XX 11-APR-1990; 90DE-4011751.
XX
XX (CONE ) CONSORT ELEKTROCHEM IND.
XX
XX Hartl P, Brem G;
XX
XX WPI; 1991-304858/42.
XX
XX
XX Recombinant DNA constructs for expressing protein in milk -
XX contg. specific mammary gland transcription control region and
XX signal sequence, providing high yield and easy prod. recovery
XX
XX Disclosure; Page 21; 41pp; German.
XX
XX
XX The rennin gene was isolated using the probe represented in AAQ14775.
XX It was used as heterologous peptide/protein together with parts of the
XX alpha-S1-casein gene in the prodn. of DNA constructs. The heterologous
XX peptide or protein may also be human insulin-like growth factor I.
XX The constructs provide high yields of the protein with simple recovery
XX from the milk. Activation of the gene occurs only in the mammary gland.
XX See also AAQ14050, AAQ14774-77.
XX
XX
XX Sequence 1210 BP; 278 A; 356 C; 324 G; 252 T; 0 other;
XX
XX
XX Query Match 4.2%; Score 49; DB 12; Length 1210;
XX Best Local Similarity 100.0%; Pred. No. 1e-14;
XX Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 656 tgtctcggttacatgacaggaatgcccagagagacatgctcagct 704
XX :||||:||||:||||:||||:||||:||||:||||:||||:
XX Db 643 tgtctcggttacatgacaggaatgcccagagagacatgctcagct 691
XX
XX
XX RESULT 8
XX ID AAT03006 standard; DNA: 1240 BP.
XX
XX AC AAT03006;
XX
XX 13-JUN-1996 (first entry)
XX
XX Chymosin open reading frame.
XX
XX
XX al-3; albino mutant; light-regulated; Neurospora; bread mould;
XX heterologous gene; expression; control; chymosin; ss.
XX
XX Mammalian sp.
XX
XX WO9530739-A1.
XX
XX 16-NOV-1995.
XX
XX 09-MAY-1995; 95MO-US05716.
XX
XX 10-MAY-1994; 94US-0240372.
XX

```

XX (UYHA-) UNIV HAWAII.
XX Kato EK, Stuart WD;
XX WPI; 1995-404108/51.
XX Nucleic acid for expression of heterologous protein - contains
PT albino promoter for light induced expression in filamentous fungi
PS Example 2; Fig 7; 26pp; English.
XX The DNA is that of a mammalian gene (open reading frame) encoding
CC chymosin. The gene was placed in operable linkage with the al-3
CC promoter (see AAT03005). The al-3 gene controls the production of
CC geranyl geranyl pyrophosphatase (GGPP) synthetase. GGPP is a precursor
CC for carotenoids and xanthophylls. It has been shown that exposure to
CC light increases the transcription level of GGPP synthetase 15-45 fold.
CC Light activates a number of genes in the common bread mould, Neurospora.
CC This can be used to regulate the expression of genes encoding
CC heterologous proteins, e.g. chymosin, in recombinant production systems.
CC Use of a light-regulated promoter is a simple and effective way to
CC control expression and allows timing to be adapted to the physiological
CC status of the host.
XX
SQ Sequence 1240 BP; 274 A; 374 C; 339 G; 253 T; 0 other;

Query Match 4.2%; Score 49; DB 16; Length 1240;
Best Local Similarity 100.0%; Pred. No. 1e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 656 tttctcgtttacatgacaggaatggccagagagatgctcacgct 704
|||||
Db 700 tttctcgtttacatgacaggaatggccagagagatgctcacgct 748

RESULT 9
AAN40055
ID AAN40055 standard; DNA; 1278 BP.
XX
AC AAN40055;
XX
DT 02-FEB-1992 (first entry)
XX
DE Sequence of prochymosin gene.
XX
KM Prochymosin expression vector; E.coli trp operon; chymosin; ss.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT CDS 3..1130
FT polyA_signal /*tag= a
FT 1245..1250 /*tag= b
XX
PN EP12175-A.
XX
PD 17-OCT-1984.
XX
PE 07-MAR-1984; 84EP-0102451.
XX
PR 09-MAR-1983; 83JP-0038439.
XX
PA (BEPF/) BEPPU T.
XX
PI Beppu T, Dozum T, Nishimori K, Shimizu N, Kawaguchi Y;
XX Hidaka M;
XX WPI; 1984-258001/42.
XX P-PSDB; AAP40078.
XX

PT Expression plasmid comprising prochymosin gene and vector -
PT useful for transforming Escherichia coli for prochymosin prodn.
XX
XX
PS Disclosure; Fig 1; 59pp; English.
XX

CC The inventors claim the prochymosin gene comprising a nucleotide
CC sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or
CC (b) the 5th codon (CGG) to the 365th codon (ATC); and recombinant
CC plasmids harboured by Escherichia coli strains deposited as FERM BP-
CC 262, -263 and -264. Any portion of the nucleotide sequence as
CC described in AAN40055 can be used. Also claimed is a vector derived
CC from plasmid pBR322. Typically plasmid pCR301 is obtd. from pOCT 2.
CC The transcriptional direction of pOCT 3 is opposite to that of
CC pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter
CC clockwise in pOCT 3.

SQ Sequence 1278 BP; 309 A; 384 C; 326 G; 259 T; 0 other;

Query Match 4.2%; Score 49; DB 5; Length 1278;
Best Local Similarity 100.0%; Pred. No. 1e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 656 tttctcgtttacatgacaggaatggccagagagatgctcacgct 704
|||||
Db 613 tttctcgtttacatgacaggaatggccagagagatgctcacgct 661

RESULT 10
AAN30022
ID AAN30022 standard; DNA; 1289 BP.
XX
AC AAN30022;
XX
DT 25-APR-1992 (first entry)
XX
DE Sequence of veal chymosin gene.
XX
KM Proteolytic enzyme; zymogen; rennin; chymosin; cheese making;
KM microbial vector; ss.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT CDS 1..1140
FT /*tag= a

XX
PN BE897201-A.
XX
PD 03-NOV-1983.
XX
PE 30-JUN-1983; 83BE-0017731.
XX
PR 13-APR-1983; 83US-0484539.
XX
PR 01-JUL-1982; 82US-0394433.
XX
PA (GENE-) GENEX CORP.
XX
DR WPI; 1983-820813/47.
XX
DR P-PSDB; AAP30013.
XX
PT Isolated chymosin or rennin and prochymosin genes - plasmid(s)
PT which replicate in prokaryotic organisms, esp. Escherichia coli,
PT and organisms used for chymosin biosynthesis
XX
XX Claim 7; Page 33-36; 43pp; French.
XX
XX The inventors claim isolated chymosin (rennin) and prochymosin genes
XX from calves, and plasmids contg. the genes which are capable of
XX replicating in a prokaryotic organism. The prokaryotic organism is
XX pref. an Escherichia species, esp. E. coli p Gx 1225 (NRRL B-15061).
XX The microorganisms transformed by the plasmid are also claimed.
XX

SO Sequence 1289 BP; 310 A; 391 C; 327 G; 260 T; 1 other;

Query Match

Best Local Similarity 4.2%; Score 49; DB 4; Length 1289;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tttctcgtttacatgacaggaatggccagagacatgtctacgct 704
|||||
Db 623 tttctcgtttacatgacaggaatggccagagacatgtctacgct 671

RESULT 11

AAN91157
ID AAN91157 standard; DNA; 1291 BP.

XX AAN91157;

DT 07-JUN-1990 (first entry)

XX Cloned sequence of (pro)chymosin.

DE Lactic acid bacteria; cheese; Streptococcus cremoris SK112; chymosin;

KM proteinase; PSK112; ss.

XX Key Location/Qualifiers

FT precursor_RNA 29..1123

FT /*tag= a

FT /product=;prochymosin

XX NL8701378-A.

XX 02-JAN-1989.

XX 12-JUN-1987; 87NL-0001378.

XX 12-JUN-1987; 87NL-0001378.

XX (NEZU-) NEDERL INS ZUIVELON.

XX PI Simons AFM, De Vos WM.

XX WPI: 1989-030097/04.

DR P-PSDB; AAP94144.

XX DNA fragment having region specific for lactic acid bacteria -

PT is contained in plasmid in microorganism used in prodn. of

XX protein and food prodn. eg cheese.

PS Disclosure; fig 2; 43pp; Dutch.

XX The DNA encoding prochymosin can be cloned into a plasmid (esp. from

CC S. cremoris SK112) and used to produce large amts of the protein by

CC recombinant DNA techniques. This could overcome the shortage of

CC prochymosin due to shortage of calf stomachs and increasing cheese

CC prodn. Prochymosin is also used in prodn. of yoghurt, butter and

CC buttermilk. See also AAN91158-N91160.

XX Sequence 1291 BP; 323 A; 385 C; 326 G; 257 T; 0 other;

QY 656 tttctcgtttacatgacaggaatggccagagacatgtctacgct 704
|||||

Db 609 tttctcgtttacatgacaggaatggccagagacatgtctacgct 657

RESULT 12

AAN30049
ID AAN30049 standard; cDNA; 1314 BP.

XX AAN30049;

DT 14-JUN-1992 (first entry)

XX cDNA sequence corresponding to one of the allelic forms (B) of

DE bovine preprochymosin.

XX Chymosin; enzyme; rennet; cheese; ss.

OS Bos taurus.

XX Key Location/Qualifiers

FT sig_peptide 24..40

FT /*tag= a

FT mat_peptide 41..1169

FT /*tag= b

XX EP77109-A.

XX 20-APR-1983.

XX 13-OCT-1982; 82EP-0201272.

XX 14-OCT-1981; 81GB-0031004.

XX (UNIL) UNILEVER NV.

XX Maat J, Verrips CT, Ledebor AM, Edens L;

XX WPI: 1983-39656K/17.

XX P-PSDB; AAP30086.

XX DNA molecules comprising genes for preprochymosin - used to

XX transform microorganisms to give strain producing the

XX prepro-enzyme and its allelic and maturation forms

XX Claim 2; Fig 1; 53pp; English.

XX Preprochymosin is an intermediate (via prochymosin and

XX pseudochymosin) for the enzyme chymosin, which is the essential milk-

XX clotting component of rennet and is used in cheese manufacture.

XX AAN30049 corresp. to mRNA isolated from the fourth stomach of a

XX preunlulant calf (abomasum, Frisian cow).

XX Sequence 1314 BP; 309 A; 398 C; 338 G; 269 T; 0 other;

RESULT 13

AAN20043
ID AAN20043 standard; DNA; 1460 BP.

XX AAN20043;

DT 16-DEC-1992 (first entry)

XX Pre-prorennin-A gene DNA sequence.

XX Pre-pro-rennin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;

XX protease; milk-clotting enzyme; ss.

OS Bos taurus.

XX Key Location/Qualifiers

CC promoters, terminators and a G418 resistance marker fused to a ADHI
CC promoter from *S.cerevisiae*.
CC The Kluyveromyces expression systems provide highly efficient secretion
CC and processing of a wide variety of proteins.
CC Sequences identical to those published in EP301669.
XX
SQ Sequence 2727 BP; 752 A; 654 C; 572 G; 749 T; 0 other;

Query Match 4.2%; Score 49; DB 10; Length 2727;
Best local similarity 100.0%; Pred. NO. 9.8e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 656 tgttctcggttacatgacaggaatggccagagagacatgctcacgct 704
|||||
Db 1268 tgttctcggttacatgacaggaatggccagagagacatgctcacgct 1316

Search completed: August 1, 2002, 01:57:49
Job time: 16964 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 18:19:35 ; Search time 5849.81 seconds
(without alignments)
2706.400 Million cell updates/sec

Title: US-09-643-755b-1

Perfect score: 1173

Sequence: 1 atgaacttccttaagcttt.....ggctagctaagcaatcgtg 1173

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	4.2	383	10	BG938086 1Ab011A08
2	49	4.2	430	10	BG937697 1Ab005E01
3	49	4.2	472	10	BG938320 1Ab015E12
4	43	3.7	399	10	BG937723 1Ab005D06
5	43	3.7	325	9	AV532492
6	23	2.0	339	10	T20903
7	23	2.0	380	9	AV538262
8	23	2.0	391	9	AA969042
9	23	2.0	404	9	AA411567
10	23	2.0	408	9	AV518232
11	23	2.0	411	9	AA411566
12	23	2.0	438	9	AA411566
13	23	2.0	483	9	AV536021
14	23	2.0	505	9	AV518638
15	23	2.0	511	9	AV518638
16	23	2.0	532	9	AV548208
17	23	2.0	549	9	AV557282

c 18	23	2.0	553	9	AV555646
c 19	23	2.0	565	9	AV532027
c 20	23	2.0	574	9	AV542531
c 21	23	2.0	581	9	AV439574
c 22	23	2.0	581	9	AV542248
c 23	23	2.0	640	9	AV537387
c 24	23	2.0	710	9	BE037991
c 25	22	1.9	153	10	BE671871
c 26	22	1.9	341	12	AQ084497
c 27	21	1.8	622	9	AV734647
c 28	21	1.8	707	9	AL668938
c 29	21	1.8	842	12	AQ896484
c 30	21	1.8	1035	12	AG151623
c 31	20	1.7	257	12	BH093689
c 32	20	1.7	405	9	AL667069
c 33	20	1.7	531	10	BE353509
c 34	20	1.7	638	9	AV875590
c 35	20	1.7	807	9	AL664431
c 36	20	1.7	1101	12	CNS050HC
c 37	19	1.6	211	9	AW754301
c 38	19	1.6	235	10	N27368
c 39	19	1.6	253	9	AV856918
c 40	19	1.6	284	9	AW749196
c 41	19	1.6	298	9	AW754326
c 42	19	1.6	301	9	AW754327
c 43	19	1.6	308	9	AT282510
c 44	19	1.6	311	9	AW858312
c 45	19	1.6	312	9	AW858289

ALIGNMENTS

RESULT 1
BG938086 383 bp mRNA linear EST 11-JUN-2001
LOCUS 1Ab011A08 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
DEFINITION
ACCESSION BG938086
VERSION BG938086.1 GI:14337458
KEYWORDS
SOURCE
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Boroidea;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
TITLE cdna's from bovine abomasum tissue
JOURNAL Unpublished (2001)
COMMENT
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
The sequence best matches gb:BPJ19786 (Bos primigenius prothymosin
mRNA, complete cds) in main database at high score of 735.0 and
E-value of 0.0
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 383
POLYA-No.

FEATURES

source
1. 383
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bovine Abomasum cDNA Library"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"

```
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMF"-strain"
/note="Organ: Abomasum; Vector: Uni-2ZAPXR; Site_1: EcoR
I; Site_2: Xho I"
BASE COUNT      83 a      111 c      113 g      76 t
ORIGIN

Query Match      4.2%; Score 49; DB 10; Length 383;
Best Local Similarity 100.0%; Pred. No. 4.2e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Yy 656 tttctcggttacctgacaggaatgcccagagagatgtctacgct 704
|||||
Db 58 TGTTCCTCGTTTACATGACAGCATGCCCGCAGAGCATGCTCAGCCT 106

RESULT 2
BG937697      430 bp      mRNA      linear      EST 11-JUN-2001
LOCUS      1AB005E01 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
DEFINITION      sequence.
ACCESSION      BG937697
VERSION      BG937697.1 GI:14337069
KEYWORDS      EST.
SOURCE      cow.
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 430)
Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
cDNA's from bovine abomasum tissue
Unpublished (2001)
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For. Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin)
mrna) in main database at high score of 844.0 and E-value of 0.0
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 430
POLYA-No.

FEATURES
source      Location/Qualifiers
1..430
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_id="Bovine Abomasum cDNA Library"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMF"-strain"
/note="Organ: Abomasum; Vector: Uni-2ZAPXR; Site_1: EcoR
I; Site_2: Xho I"
BASE COUNT      97 a      130 c      119 g      84 t
ORIGIN

Query Match      4.2%; Score 49; DB 10; Length 430;
Best Local Similarity 100.0%; Pred. No. 4.3e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Yy 656 tttctcggttacctgacaggaatgcccagagagatgtctacgct 704
|||||
Db 52 TGTTCCTCGTTTACATGACAGCATGCCCGCAGAGCATGCTCAGCCT 100
```

```
RESULT 3
BG938320      472 bp      mRNA      linear      EST 11-JUN-2001
LOCUS      1AB015E12 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
DEFINITION      sequence.
ACCESSION      BG938320
VERSION      BG938320.1 GI:14337692
KEYWORDS      EST.
SOURCE      cow.
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 472)
Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
cDNA's from bovine abomasum tissue
Unpublished (2001)
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For. Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin)
mrna) in main database at high score of 928.0 and E-value of 0.0
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 472
POLYA-No.

FEATURES
source      Location/Qualifiers
1..472
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_id="Bovine Abomasum cDNA Library"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMF"-strain"
/note="Organ: Abomasum; Vector: Uni-2ZAPXR; Site_1: EcoR
I; Site_2: Xho I"
BASE COUNT      104 a      143 c      131 g      94 t
ORIGIN

Query Match      4.2%; Score 49; DB 10; Length 472;
Best Local Similarity 100.0%; Pred. No. 4.4e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Yy 656 tttctcggttacctgacaggaatgcccagagagatgtctacgct 704
|||||
Db 65 TGTTCCTCGTTTACATGACAGCATGCCCGCAGAGCATGCTCAGCCT 113

RESULT 4
BG937723      399 bp      mRNA      linear      EST 11-JUN-2001
LOCUS      1AB005D06 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
DEFINITION      sequence.
ACCESSION      BG937723
VERSION      BG937723.1 GI:14337095
KEYWORDS      EST.
SOURCE      cow.
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 399)
```

AUTHORS Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
TITLE CDNA's from bovine abomasum tissue
JOURNAL Unpublished (2001)
COMMENT Contact: Dr. Stephen Moore
. Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca

The sequence best matches gb:BP019786 (Bos primigenius prothymosin mRNA, complete cds) in main database at high score of 767.0 and E-value of 0.0

PCR Primers

FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: 73 primer
High quality sequence stop: 399
POLYA-No.

FEATURES

Source

1..399
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bovine Abomasum cDNA library"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMRF", strain"
/note="Organ: Abomasum; Vector: Uni-2ZAPYK; Site_1: Ecor
I; Site_2: Xho I"
89 a 121 c 112 g 77 t

BASE COUNT

89 a 121 c 112 g 77 t

ORIGIN

Query Match 3.7%; Score 43; DB 10; Length 399;
Best Local Similarity 100.0%; Pred. No. 5.3e-11;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CGGTTTACATGACAGAGATGGCAGAGATGCTCAGCT 43

662 cggttacatgacagagatggcagagatgctcagct 704

|||||

RESULT 5

LOCUS

AV532492 325 bp mRNA linear EST 01-SEP-2000

DEFINITION

AV532492 Arabidopsis thaliana flower buds Columbia Arabidopsis
thaliana cDNA clone FB043b03f 3', mRNA sequence.

ACCESSION

AV532492

KEYWORDS

EST.

SOURCE

thale cress.

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
1 (bases 1 to 325)

REFERENCE
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries

JOURNAL

MEDLINE

20363093

COMMENT

Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers

FEATURES

Source

1..325
/organism="Arabidopsis thaliana"
/strain="Columbia"

/db_xref="taxon:3702"
/clone_lib="FB043b03f"
/clone_lib="Arabidopsis thaliana flower buds Columbia"
/tissue_type="flower buds"
/note="Vector: pBluescriptII SK-; Site_1: EcorI; Site_2:
XhoI"

BASE COUNT 106 a 92 c 59 g 68 t

ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 325;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 209 TGGATCTTGGAGATGTGTCAT 187

1087 TGGATCTTGGAGATGTGTCAT 1109

ORIGIN

RESULT 6

LOCUS

T20903 339 bp mRNA linear EST 07-JAN-1998
DEFINITION 2911 Lambda-PRL2 Arabidopsis thaliana cDNA clone 89F977, mRNA
sequence.

ACCESSION

VERSION

T20903.1 GI:2756823

KEYWORDS

SOURCE

thale cress.

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
1 (bases 1 to 339)

REFERENCE
AUTHORS Newman,T., deBrujn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
N., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
E. and Somerville,C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones

Plant Physiol. 106, 1241-1255 (1994)
9518729
On Jan 7, 1998 this sequence version replaced gi:502344.
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313t@msu.edu
Seq primer: 77.

JOURNAL

9518729

MEDLINE

COMMENT

Location/Qualifiers
1..339
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone_lib="89F977"
/clone_lib="Lambda-PRL2"
/note="Vector: lambda ZAP-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques). The vector is BRL's lambda ZAP-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA.

FEATURES

Source

1..339

/organism="Arabidopsis thaliana"

/strain="var columbia"

/db_xref="taxon:3702"

/clone_lib="89F977"

/clone_lib="Lambda-PRL2"

/note="Vector: lambda ZAP-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques). The vector is BRL's lambda ZAP-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA.

BASE COUNT 67 a 61 c 92 g 106 t 13 others

ORIGIN

Query Match 2.0%; Score 23; DB 10; Length 339;
Best Local Similarity 100.0%; Pred. No. 1.1;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 tggatcttggagatgtgtcat 1109
 |||
 Db 82 TGGATCTTGGGAGATGTGTTCA 104

RESULT 7
 AV538262 380 bp mRNA linear EST 06-SEP-2000
 LOCUS AV538262 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
 DEFINITION CDNA clone R2113a03f 3', mRNA sequence.

ACCESSION AV538262
 VERSION AV538262.1 GI:8698545

KEYWORDS EST.
 SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE Asariyota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

1 (bases 1 to 380)
 Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S.
 DNA Res. 7, 175-180 (2000)

JOURNAL 20363093
 MEDLINE Contact: Erika Asamizu
 COMMENT The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute
 Yana 1532-3, Kasarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 Location/Qualifiers

FEATURES
 source
 1..380
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="R2113a03f"
 /clone_1lb="Arabidopsis thaliana roots Columbia"
 /tissue_type="roots"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 730 a 96 c 65 g 89 t

ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 tggatcttggagatgtgtcat 1109
 |||
 Db 332 TGGATCTTGGGAGATGTGTTCA 310

RESULT 8
 AA969042 391 bp mRNA linear EST 07-JUL-1998
 LOCUS AA969042 op3d07.s1 Soares.NFL.T.GBC.S1 Homo sapiens CDNA clone
 DEFINITION IMAGE:1579597 3' similar to TR:Q27951 PROCHYMOSIN.; mRNA
 sequence.

ACCESSION AA969042
 VERSION AA969042.1 GI:3144222

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 391)
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 619 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 378.
 Location/Qualifiers

FEATURES

source

1..391
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1579597"
 /clone_1lb="Soares.NFL.T.GBC.S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pUT30-Pac (Pharmacia) with
 a modified polylinker. Site_1: Not I; Site_2: Eco RI;
 equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NbHL19w, testis NHT, and B-cell
 NCI-CCGAP GCBI) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 1.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 82 a 101 c 115 g 93 t

ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1015 cccctcgccataccagcagga 1037
 |||
 Db 256 CCTCCGCGCTATACGACGAGA 234

RESULT 9
 AA411567 404 bp mRNA linear EST 17-MAY-1997
 LOCUS AA411567 zv22901.s1 Soares.NHMPU.S1 Homo sapiens CDNA clone IMAGE:75416 3'
 DEFINITION similar to SW:CHYM_SHEEP P18276 PROCHYMOSIN PRECURSOR, B-FORM;;
 mRNA sequence.

ACCESSION AA411567
 VERSION AA411567.1 GI:2069151

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 404)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Maria, M., Martin, D., Moore, B.,
 Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie,
 T., Waterston, R., and Wilson, R.

Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estevatson.wustl.edu

Unpublished (1997)
 Contact: Wilson RK

This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Possible reversed clone: similarity on wrong strand

Seq primer: -41m13 fwd. ET from Amersham
 High quality sequence stop: 374.
 Location/Qualifiers

FEATURES

source

1..404
 /organism="Homo sapiens"

/db_xref="GDB:5977315"
/db_xref="taxon:9606"
/clone_image="754416"
/clone_lib="Soares.NhHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBH, pregnant uterus NBHPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 88 a 105 c 116 g 95 t

ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1015 cccctcgctaccagccagga 1037
|||||
Db 257 cccctcgctaccagccagga 235

RESULT 10
AV518232 408 bp mRNA linear EST 06-SEP-2000
LOCUS AV518232 Arabidopsis thaliana aboveground organs two to six-week
DEFINITION Old Arabidopsis thaliana cDNA clone APD1301F 3', mRNA sequence.
ACCESSION AV518232
VERSION AV518232.1 GI:8677759
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 408)
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. 408
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_image="APD1301F"
/clone_lib="Arabidopsis thaliana aboveground organs two to six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 135 a 112 c 69 g 92 t

ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 408;

Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1087 tggatctggagatgtgtcat 1109
|||||
Db 285 TGGATCTGGAGATGTGTTCAT 263

RESULT 11
AV559941/c 411 bp mRNA linear EST 07-SEP-2000
LOCUS AV559941 Arabidopsis thaliana green siliques Columbia Arabidopsis
DEFINITION thaliana cDNA clone SQ126C01F 3', mRNA sequence.
ACCESSION AV559941
VERSION AV559941.1 GI:8731367
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 411)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. 411
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_image="SQ126C01F"
/clone_lib="Arabidopsis thaliana green siliques"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 136 a 109 c 70 g 96 t

ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1087 tggatctggagatgtgtcat 1109
|||||
Db 332 TGGATCTGGAGATGTGTTCAT 310

RESULT 12
AA411566 438 bp mRNA linear EST 17-MAY-1997
LOCUS AA411566 zw22g01.f1 Soares.NhHMPu_S1 Homo sapiens cDNA IMAGE:754416 5'
DEFINITION similar to TR:G457097 G457097 CHYMOSIN C. [1] ;, mRNA sequence.
ACCESSION AA411566
VERSION AA411566.1 GI:2069150
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 438)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.

TITLE Washu-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28m3 rev2. ET from Amer sham.

FEATURES
source
1. 438
/organism="Homo sapiens"
/db_xref="GDB:5977315"
/db_xref="taxon:9606"
/clone_lib="IMAGE:754416"
/clone_1lb="Soares_NbHMPU_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker. Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NDH, pregnant uterus
NbHPU, and fetal heart NbHHL19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 103 a 124 c 117 g 94 t
ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1015 cccctcgctataccagcagga 1037
|||||
Db 187 CCCCGCGCTATACCGACGACGA 209

RESULT 13
AV536021/c 483 bp mRNA linear EST 06-SEP-2000
LOCUS AV536021 Arabidopsis thaliana flower buds Columbia Arabidopsis
DEFINITION thaliana cDNA clone FB106C09F 3', mRNA sequence.
ACCESSION AV536021
VERSION AV536021.1 GI:8696304
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (phases 1 to 483)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093

JOURNAL MEDLINE
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1. 483
/organism="Arabidopsis thaliana"

/strain="Columbia"
/db_xref="taxon:3702"
/clone="FB106C09F"
/clone_1lb="Arabidopsis thaliana flower buds Columbia"
/tissue_type="flower buds"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 161 a 123 c 85 g 114 t
ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 483;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1087 tggatctggagatgcttcat 1109
|||||
Db 338 TGGATCTTGGAGATGCTTCAT 316

RESULT 14
A1479358/c 505 bp mRNA linear EST 14-APR-1999
LOCUS A1479358
DEFINITION tm27607.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2157828 3' similar to TR:028950 Q28950 PREPROCHYMOSIN
PRECURSOR ; mRNA sequence.
ACCESSION A1479358
VERSION A1479358.1 GI:4372526
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (phases 1 to 505)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-rt@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 608 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 367.
Location/Qualifiers

FEATURES
source
1. 505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2157828"
/clone_1lb="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker. Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHHL19W, testis NHT, and B-cell
NbHHL19W) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.

BASE COUNT 106 a 138 c 141 g 119 t 1 others
ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1015 cccctcgctataccagcagga 1037

Db 271 CCCFCCGCTATACGACGACAGA 249

RESULT 15

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

Search completed: July 31, 2002, 23:07:30
job time: 17275 sec

AV518638 511 bp mRNA linear EST 06-SEP-2000

AV518638 Arabidopsis thaliana aboveground organs two to six-week

old Arabidopsis thaliana cDNA clone APD34e06f 3', mRNA sequence.

AV518638.1 GI:8678165

EST.

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 511)

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

A large scale analysis of cDNA in Arabidopsis thaliana: Generation

of 12,028 non-redundant expressed sequence tags from normalized and

size-selected cDNA libraries

DNA Res. 7, 175-180 (2000)

20363093

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizukazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

Location/Qualifiers

1. 511

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="APD34e06f"

/clone_lib="Arabidopsis thaliana aboveground organs two to

six-week old"

/tissue_type="aboveground organs"

/dev_stage="two to six-week old"

/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:

XhoI"

156 a 135 c 102 g 118 t

2.0%; Score 23; DB 9; Length 511;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1087 tggatctggagatgttcat 1109

|||||

274 TGGATCTGGAGATGTTCAT 252

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2002, 01:41:47 ; Search time 9238.6 Seconds
(without alignments)
8963.079 Million cell updates/sec

Title: US-09-643-755B-3
3957
Sequence: 1 ctgcaggaattcatgtact.....acctcaccactaagttacc 3957

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pi:*
25: em_pi:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_higo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID Description

1	3957	100.0	3957	6	AX088021	AX088021 Sequence
2	1551	39.2	3502	6	PVBCSP	J01263 Phaeosolus v
3	1173	29.6	1173	6	AX088019	AX088019 Sequence
4	1047	26.5	1415	6	AX343913	AX343913 Sequence
5	887	22.4	2970	6	AX252300	AX252300 Sequence
6	181	4.6	1990	6	E00452	E00452 DNA sequence
7	181	4.6	2090	6	A06496	A06496 Artificial
8	169	4.3	4764	6	PVAPHASE	X52626 Phaeosolus v
9	133	3.4	1475	6	A11822	A11822 Artificial
10	133	3.4	1475	6	PVPHASBR	X03004 Phaeosolus v
11	95	2.4	207	8	ZMEINI	X06175 Zea mays/P.
12	90	2.3	1535	8	PVPHASAR	X02960 Phaeosolus v
13	77	1.9	1454	8	U01131	U01131 Phaeosolus v
14	75	1.9	348	8	ZMEINI2	X06176 Maize chime
15	63	1.6	1478	8	U01121	U01121 Phaeosolus l
16	57	1.4	1430	8	U01132	U01132 Phaeosolus v
17	56	1.4	1268	4	BOVCHYMOC	J00004 Bovine chym
18	49	1.2	1098	6	E00079	E00079 CDNA sequen
19	49	1.2	1098	6	E02341	E02341 CDNA sequen
20	49	1.2	1101	4	AF177290	AF177290 Bubalus a
21	49	1.2	1115	12	SYNPROCA	M22593 Synthetic b
22	49	1.2	1175	6	I04058	I04058 Sequence 5
23	49	1.2	1240	6	AR002347	AR002347 Sequence
24	49	1.2	1275	4	BOVCHYMOA	J00002 bovine chym
25	49	1.2	1289	6	E00144	E00144 CDNA encodi
26	49	1.2	1291	6	A15836	A15836 chymosin ge
27	49	1.2	1305	4	BOVCHYMOB	J00003 bovine chym
28	49	1.2	1460	6	E00042	E00042 DNA coding
29	49	1.2	1460	6	E00295	E00295 CDNA encodi
30	49	1.2	2726	6	I08097	I08097 Sequence 5
31	49	1.2	2733	6	AR073077	AR073077 Sequence
32	49	1.2	2982	6	I08098	I08098 Sequence 1
33	44	1.1	1094	4	BP019786	U19786 Bos primige
34	38	1.0	1290	6	E00075	E00075 CDNA encodi
35	38	1.0	1291	6	A15633	A15633 preprochymo
36	37	0.9	1311	6	E00108	E00108 DNA coding
37	37	0.9	119	4	BOVCHYMO8	M14076 Bovine chym
38	37	0.9	1404	8	U01122	U01122 Phaeosolus l
39	35	0.9	171	4	BOVCHYMO2	M14070 Bovine chym
40	35	0.9	637	6	E05472	E05472 DNA sequenc
41	33	0.8	179	6	I04059	I04059 Sequence 6
42	33	0.8	1172	4	CDR131677	AJ131677 Camelus d
43	32	0.8	335	4	BOVCHYMO9	M14077 Bovine chym
44	32	0.8	1292	4	OAPPCHY	X53037 Lamb mRNA f
45	30	0.8	134	4	BOVCHYMO6	M14074 Bovine chym

ALIGNMENTS

RESULT 1
AX088021 LOCUS AX088021 3957 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 3 from Patent WO0114571.
ACCESSION AX088021
VERSION AX088021.1 GI:13396949
KEYWORDS

SOURCE
ORGANISM synthetic construct.
synthetic construct.
artificial sequence.
1 (bases 1 to 3957)
REFERENCE
AUTHORS van Rooijen,G., Keon,R.G., Boothe,J. and Shen,Y.
TITLE Commercial production of chymosin in plants
JOURNAL Patent: WO 0114571-A 3 01-MAR-2001;
Sembiosys Genetics Inc. (CA)
FEATURES
source location/Qualifiers
1..3957
/organism="synthetic construct"
/db_xref="taxon:32630"
1554..2726
/note="Figure 2"
/codon_start=1
/transl_table=11

CDS

|||||
Db 1921 AAAATTTCGATCCGAGAAAGCGTCACCTTCGAACTTAGGCAAAACCTTGTCTATAC 1980
Oy 1981 actacggtacaggtacgcatcgaagaaatcttaggtatgtatgacgtatctctccaca 2040
Db 1981 ACTACGGTACAGGTACGTAGTGAAGAACTTAGCTATGATACCGTCACTCTCCAAACA 2040
Oy 2041 ttgttgacatccaacagacagtaagacttagcacccaagaaaccagtgatgtcttcacct 2100
Db 2041 TTGTGGACATTTCAACACAGTAGTACTTAGCACCCAAGAACAGGTGATGTCTTCACCT 2100
Oy 2101 atgcagaatctcgatgtgcatcccttggtatggaatacccatcgctcgctcagaagtactga 2160
Db 2101 ATGCAGATTCGATGCGCATCTTGTGTATGGCATCCCATCCCTCGCGCAGAGTACTGCA 2160
Oy 2161 taccgtgtttgacacatagatagacagcaaccagtaagctcaagactgtgttcggttt 2220
Db 2161 TACCTGTGTTGACAAATAGTAGTAACCGACACCTAGTAGCTCAAGACTGTGTTCTCGGTTT 2220
Oy 2221 acatgagcaggaatgagccaggaagacatgctcaacgcttgagcatctgacatccact 2280
Db 2221 ACATGAGACAGGAATGGCAGAGAGACATGCTCAGCGTTGGAGCTATGTATCCATCTACT 2280
Oy 2281 acacaggaatctctcactggttccagtcacatgtgcagcagtaactggaatctcgtg 2340
Db 2281 ACACAGGATCTTCTCACTGGGTTCCAGTCACTGTGACAGCACTACTGGCAATCTACTGTG 2340
Oy 2341 acaggtgcacacatcagcaggtggtgtgtgtgcatgtgaaagtggagtgccaagtctctgg 2400
Db 2341 ACAGGTGCACATCAGGAGGTTGGTGTGTGATGTGAAGGTGATGTCAAACCTTCTTGG 2400
Oy 2401 ataccggtacgtccaaagctgtgcagacctagcagcagcatctcaacatcagaagacta 2460
Db 2401 ATACCGGTAGTCCAACTGCTGCGACCTAGCACAGCATTTCTTAACATTTCAGAACGTA 2460
Oy 2461 ttggaagccacacagaaacagtaagtgagtttgacatagatgtgcgaacaccttagctaca 2520
Db 2461 TTGGAGCCACACAGAACAGTACGCTGAGTTTGACATAGATTGGCAACACTTAGCTACA 2520
Oy 2521 tgcctaaagtgtctttgagatatacaagcgcaagatgtaccactgaaccccccgctata 2580
Db 2521 TGCTTACAGTTGCTTTGAGATCAACGGCAAGATGACCCCTGACCCCTCCCTATA 2580
Oy 2581 ccaagccagatccaagggctctgcacagtgatccagagtgcgaacatcccaagaat 2640
Db 2581 CCAAGCCAGATCCAAGGGTCTCTGCACAGTGGATTCAGAGTGAACCATTTCCAGAAAT 2640
Oy 2641 ggaatctggagaatgltgltcaatcgtgagtaactaacgctcttgacaagggccaacaac 2700
Db 2641 GGATCTTGGGAGATGTGTCTATCTGATGATCTACAGCGTCTTGAACAGGGCCAAACAC 2700
Oy 2701 tctgttggtcagctaaagcaatctgaaagcttaataagtaagtaactaaatgcatgtggt 2760
Db 2701 TCGTTGGGCTAGCTAAAGCAATCTGAAGCTTAATTAAGTAACTAAATGCAATGATAGGT 2760
Oy 2761 gtaagaagcactaagagacatgagatactgataccgacatgtaaacagtaataactga 2820
Db 2761 GTAAGACCTCATGAGAGCATGGAATTTGTATCCGACATGTAACAGTATTAATTAAGTGA 2820
Oy 2821 gctcatctcaactctctctatgaaataaacaagaagtgltatgataataaactctatc 2880
Db 2821 GCTCATCTCACTCTCTATGATTAACAAGAGATGTATGATATTAACACTCTATC 2880
Oy 2881 tatgacctatgtctatagataaattctcttatatatataatcatctgaatgtg 2940
Db 2881 TATGACCTTATATGTTATATGATTAATTTCTCTTATTAATTAATCATCTGATGATGTG 2940
Oy 2941 acggtcatgaaatgctccaatagtaacaaaacaatgltgtactataagactctcaaa 3000
Db 2941 ACGGCTTATGGAATGCTTCAAAATAGTACAAAACAATGTTACTATTAAGACTTTTCAAA 3000
Oy 3001 caattctaacttaagcatgttgaaagacataaagtglttaagaacataaactataa 3060
|||||

Db 3001 CAATTCTAATTTAGCATGTGAACGAGACATAGGTGTAAGAACATTAATTAATA 3060
Oy 3061 tggagaaggttgcctccattatataatattaccactatgatatattaga 3120
Db 3061 TGGAGAAGGTGCTCCATTATATATATATATATATATATATATATATATATATAGGA 3120
Oy 3121 tgttaaggaagacatacaatataaagagaagtttgatccatltatataatataac 3180
Db 3121 TGTTAAGSAGACATTAACATTAATAAGAGAAAGTTGTATCCATTATATATATATATAC 3180
Oy 3181 taccattatataataactataccactatctaagtccttataaagttgatccatg 3240
Db 3181 TACCATTTATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3240
Oy 3241 atattcctaattttagttgatatgataagaaggtaactattggaacctctactc 3300
Db 3241 ATATTCTTAATATTTTAT 3300
Oy 3301 tglataaaggttgatcatcccttaagtggttctatataattatgtctctacagat 3360
Db 3301 TGTATTAAGGTGGATCATCTCTTAAGTGGGTCTATTTATTTATTTCTCTTACAGAT 3360
Oy 3361 aaaaaaaaatlatgagltggttgcataaaatlatgaaagalttaaaataaataa 3420
Db 3361 AAAAAAAAATTAATGAGTTGGTTGATTAATTAATTAATTAATTAATTAATTAATTAAT 3420
Oy 3421 ataataacataataatataatataaattataataataataataataataataa 3480
Db 3421 ATAAATTAACAT 3480
Oy 3481 agtaataatgltcataaatacaataacgtlttaagccttgtagcagcctcaatatt 3540
Db 3481 AGTAATATTTCTCAATTAATATATATATATATATATATATATATATATATATATAT 3540
Oy 3541 taacgagagtaaaacataatgtgacttltgttalttbaacaatataataataactat 3600
Db 3541 TAAACGAGAGTAACATATTTGACTTTTGGTATTTTAACAAATTAATTAATTAACACTAT 3600
Oy 3601 atgaatattttttttatgtgcaagaaataaataataatttgaggggcaatggt 3660
Db 3601 ATGAATATTTTATTTTATATGGAAGAAATAAATAATTAATTAATTAATTAATTAAT 3660
Oy 3661 gtcccaatcctctatacaaccacttccacaggaagtgcaagtcgagggcaacaaaac 3720
Db 3661 GTCCCAATCCTTATACAAACCAACTTCCACAGAAAGTAGTGGGCAACCAAAAAAAC 3720
Oy 3721 aggcagaaggaatctttaaatttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3780
Db 3721 AGGCAAGGAAATTTTAAATTTGGGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 3780
Oy 3781 cactacacataacccttttagcagtagagcaatggttgaccggtgtgcttgaactcttta 3840
Db 3781 CACTACACATTAACCTTTTACAGATAGAGCAATGTTGACCGTGTGCTTACCTCTTTTA 3840
Oy 3841 tttaatttttttcaagcaagaataaataaataaataaataaataaataaataaataaataa 3900
Db 3841 TTTTATTTTATATCAGCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3900
Oy 3901 aaccctatacaaaaaccccaaaaacaaagtctctcagaacccctacaactaaagtiacc 3957
Db 3901 AACCTTATACAAAACCCCAAAAACAAGTTCTTAGCACCTTACCACTAAGTAAAGTACC 3957

RESULT 2
PHVBCSP PHVBCSP 3502 bp DNA linear PLN 17-JUN-1998
LOCUS Phaseolus vulgaris beta-type phaseolin storage protein gene,
DEFINITION complete cds.
ACCESSION J01263 M13758
VERSION J01263.1 GI:3228361
KEYWORDS
SOURCE Phaseolus vulgaris.
ORGANISM Phaseolus vulgaris

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus

1 (bases 1371 to 2450; 2466 to 3294; 3322 to 3502)

Slightom,J.L., Sun,S.S.M. and Hall,T.C.
Complete nucleotide sequence of a French bean storage protein gene: phaseolin

Proc. Natl. Acad. Sci. U.S.A. 80, 1897-1901 (1983)

2 (bases 1 to 3502)

Doyle,J.J., Schuller,M.A., Godette,W.D., Zenger,Y., Beachy,R.N. and Slightom,J.L.
The glycosylated seed storage proteins of glycine max and phaseolus vulgaris: Structural homologes of genes and proteins

J. Biol. Chem. 261, 9228-9238 (1986)

3 (bases 2921 to 3502)

Hall,T.
Direct Submission
Submitted (27-APR-1993) Biology, Texas A&M University, College Station, TX 77843-3155, USA

4 (bases 1 to 3502)

Hall,T.
Direct Submission
Submitted (16-JUN-1998) Biology, Texas A&M University, College Station, TX 77843-3155, USA

REMARK
COMMENT
FEATURES
SOURCE

Nucleotide sequence updated by submitter
On Jun 16, 1998 this sequence version replaced gi:169322.
Location/Qualifiers

1..3502

/organism="Phaseolus vulgaris"
/db_xref="taxon:3885"
/clone_1fb="AF-pvph3.8; pvph 177.4-E1.6"
/clone_1lb="Charon 24a clone 177.4"
join(1548..1860,1933..2123,2212..2292,2417..2662,
2791..3049,3153..3370)
/gene="beta-ppsp"
/gene="beta-ppsp"
/number=1
join(1548..1860,1933..2123,2212..2292,2417..2662,
2791..3049,3153..3370)
/gene="beta-ppsp"
/codon_start=1
/product="beta-type phaseolin storage protein"
/protein_id="AAC23610.1"
/db_xref="GI:169323"

/translation="MMARVPLLLGLFLASISAFATSLREEESQNPFTFNSDN
SWMTLFRKNOGHIRVLQRFDOQSKRLQNEIDYRLVERRSKPEYLLPQDADELLVY
RSGSALIVLVKPDREYFPLTSDNPIFSHQKIPAGTIIYLVNPPKEDRLIOLAM
PVNPQIHEFFLSSTEQOQYLOEFKSHLEAFNSKFEELINRVLFEEEOQEOGOE
GIVINIDSEQIKELSKHAKSSRSKSLSKONTGNEGTERTDNLANLISIEHE
EGALFVPHYSKAIVILVNEGRAHVELNPKNKETLPEESYRAELSKRDVVIIPA
YPAIKATSNVNTGFGFINANNNNRNILAKTQNVISISIRALDGKDYLLTSGSGD
EVKLLINKSGSFTVDAAHHQDEQKSGHOEQKSGKGAFFV"
1548..1619
/gene="beta-ppsp"
join(1620..1860,1933..2123,2212..2292,2417..2662,
2791..3049,3153..3367)
/gene="beta-ppsp"
/product="beta-type phaseolin storage protein"
1861..1932
/gene="beta-ppsp"
/number=1
1933..2123
/gene="beta-ppsp"
/number=2
2124..2211
/gene="beta-ppsp"
/note="putative"
/number=2
2212..2292
/gene="beta-ppsp"

exon
intron
intron
exon
mat_peptide
sig_peptide

intron
exon
intron
exon
intron
exon
intron
exon
BASE COUNT
ORIGIN

/number=3
2293..2416
/gene="beta-ppsp"
/note="putative"
/number=3
2417..2662
/gene="beta-ppsp"
/number=4
2663..2790
/gene="beta-ppsp"
/note="putative"
/number=4
2791..3049
/gene="beta-ppsp"
/number=5
3050..3152
/gene="beta-ppsp"
/note="putative"
/number=5
3153..>3370
/gene="beta-ppsp"
/number=6

1124 a 678 c 572 g 1128 t

Query Match 39.2%; Score 1551; DB 8; Length 3502;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gaattcattgtactccagatcatcatlaagtgaaagtttggctctcgcggtggtt 66
|||||
Db 1 GAATTCATTGTACTCCAGATCATCATLAAGTGAAGTTTGGCTCTCGCGGTGTT 60
QY 67 ttaccctatttaagggtgttccacctaanaattcgtgatcatcttacttact 126
|||||
Db 61 TTACCTCTATTTAAGGGGTTTCCACCTAAATCTGGATCATCTTACTTACTT 120
QY 127 gttaacttaattctcataactcttggttgaatatacagcttcgcagatatacc 186
|||||
Db 121 GTTACTTTAATTCTCATTAATCTTGGTTGAATATACCGCTCCGACGATATCC 180
QY 187 tacaatttattatctgtttaaacaatttcaaacgcataaatttlaagatccgct 246
|||||
Db 181 TACAAATTTATTATTGTTAAACATTTTCAACCGCATTAATTTTGAAGTCCGCT 240
QY 247 atctttaatgagctcaacatttccatatatgaataataataattacttaatt 306
|||||
Db 241 ATCTTTAATGTAAGTCTTAACATTTTCATATTAATAATAATAATTTTACGCT 300
QY 307 gtagaagacataaagatttattcttattcttccatataaagtttaataataat 366
|||||
Db 301 GGTGAAAGCATTAAGATTATTCTTATCTCTTCAATATAATGTTAATATCAATAT 360
QY 367 aaacaatttcttaacctaaagaagatttccatttataatttaanaatatattca 426
|||||
Db 361 AACCAATTTCTTACCTTAAGAGATTCCCATTTTATATTTAAAAATATTATATCA 420
QY 427 aatatlttcaacacgtaaatctcataataatagttgttcaaaagtaataaattta 486
|||||
Db 421 AATATTTTTCACACGTAATATCTCATTAATTAAGTTGTTCAAAAGTAATAATTTTA 480
QY 487 actcataatttlttattgactgactgttaagcaaacaccagtgacacaactagcoat 546
|||||
Db 481 ACTCATATATTTTATTTCAGCTGATCTTAAGCAACACCAAGTACACACTAGCCAT 540
QY 547 ttttcttgaataaaaaaaccaattatcatgttatttttatacaatgaanaatt 606
|||||
Db 541 TTTTTCCTTGAATAAAAAATCAATATATCATGTATTTTATATACATGAATAATTT 600
QY 607 caccacaacatcttgtgtatttcttgaagaagatcatgttatgcaaaattctataat 666
|||||
Db 601 CACCAACCAATCATTTGTGTATTCTGAAGCAAGTATGATGTAACAAATTTATATT 660

QY	667	cccatcttgacactacggaagtaactcgtgaacacctcgtcttacaatcgagagacacatctctcta	726
Db	661	cccatcttgacactacggaagtaactcgtgaacacctcgtcttacaatcgagagacacatctctcta	720
QY	727	aagtaatttaataagttactatattcaaatcttcaatataataaatactcaattactac	786
Db	721	aagtaatttttaataagttactatattcaaatcttcaatataataaatactcaattactac	780
QY	787	ttctaaaaaataatataagatacaatataaataataactttcaattcaattcaattcaattgt	846
Db	781	ttctaaaaaataatataagttactatattcaaatcttcaatataataaatactcaattcaattgt	840
QY	847	tgaattcttgactatgacttatactctactactagtttaaatctgtttatagatgltta	906
Db	841	tgaatttttgactaattgattttatt	900
QY	907	aagtaaatataagttatagttatagttatagttatagttatagttatagttatagttatagttat	966
Db	901	aagtaaatataagttatagttatagttatagttatagttatagttatagttatagttatagttat	960
QY	967	attatctgtgagactaatcttcaatataattctctatctgtcttcaactcttcttctgtatgtaa	1026
Db	961	att	1020
QY	1027	gtccgttaactagaaatcaagtggtgtgcacatgagacactcgtgtgtcttctgtttcaat	1086
Db	1021	gtccgttaactagaaatcaagtt	1080
QY	1087	gggtcttcttgagagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa	1146
Db	1081	gggtcttcttgagagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa	1140
QY	1147	gcaatcacacaacacacttcaataatagtcactgtgtatcaagaatgcgcgcgtcactgtat	1206
Db	1141	gcaatcacacaacacacacttcaataatagtcactgtgtatcaagaatgcgcgcgtcactgtat	1200
QY	1207	gtcttaattgcaatgcaagcaaacagtgctctataatgacatttaaatggtcttaacccatct	1266
Db	1201	gtcttaattgcaatgcaagcaaacagtgctctataatgacatttaaatggtcttaacccatct	1260
QY	1267	caaccac	1326
Db	1261	caaccac	1320
QY	1327	tcaattctcttcgcgaactcaattcttcttcaacttcaacacacacacacacacacacacacacac	1386
Db	1321	tcaattctcttcgc	1380
QY	1387	tcattccatgccaatctgcacatgattgttccaaacacacacacacacacacacacacacacacac	1446
Db	1381	tcattccatgccaatctgcacatgattgttccaaacacacacacacacacacacacacacacacac	1440
QY	1447	taaaatacctctataatacactcactcttcttcaatcactcaatcactcaatcactcaatcactc	1506
Db	1441	taaaatacctctataatacactcactcttcttcaatcactcaatcactcaatcactcaatcactc	1500
QY	1507	tactactataatacccccaaccccaactcatatccaataactactactactatga	1557
Db	1501	tactactataataatacccccaaccccaactcatatccaataactactactactatga	1551
RESULT 3			
LOCUS	AX088019	1173 bp	DNA
DEFINITION	Sequence 1 from Patent WO0114571.	Linear	PAT 17-MAR-2001
ACCESSION	AX088019		
VERSION	AX088019.1	GI:13396947	
KEYWORDS			
SOURCE			
ORGANISM	Bos taurus		
	Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		

REFERENCE	Bovidae; Bovinae; Bos.			
AUTHORS	1 (bases 1 to 1173)			
TITLE	van Rooijen, G., Keon, R. G., Boothe, J. and Shen, Y.			
JOURNAL	Commercial production of chymosin in plants			
FEATURES	Patent: WO 0114571-A 1 01-MAR-2001; Sembiosys Genetics Inc. (CA) Location/Qualifiers 1..1173 /organism="Bos taurus" /db_xref="taxon:9913" 1..1173 /note="unnamed protein product" /codon_start=1 /protein_id="CAC34680.1" /db_xref="GI:13396948" /translation="MNFLKSPFPFAFLFGFYEVAVTHAAETRIPLPKGSLKALK ENGLIEDFLKQOYGISKYSRGEVSAVLPTNLDSOYREK IYLGIPROEFLPDP GSSDPWVSITCKSNACKNHOREDPKRSSITONIKGLPISTHYGSGMOGLIGYTVTV SNVIDQQTGLSTQEPDGVTFVAFPGILGMAYPSLASESLVFPNNMRHLVAD LFSYVMNRNGESMLTLEALIDPSYVYTGSLHWPVYQDWFVDSYVLSGVVACGEG GCOALIDTGSTLVGSSDILNIOALGATNOGEGPIDDNLSYMPVAFVEFNGKM YPLPNSATSDOGFPGTSGFSEMHSSKWLIGDVFIREYYSVFPRRANLVGLAKAI"			
BASE COUNT	299 a	308 c	262 g	304 t
ORIGIN				
Query Match	29.6%	Score 1173:	DB 6:	Length 1173:
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1173:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1554	atgaacttccttaagctcttccttccttctaagcttccttccttcttgctcgcaatactcgt	1613	
Db	1	ATGAACCTTCCTTAAGCTTCCTTCCTTTCTACAGCTTCCTTGTTGGTGCACATACCTCGTT	60	
QY	1614	gctgttactacagctgcgtcgagataccaccgcatctcctctacaaagtgatcttcgt	1673	
Db	61	GCTGTTACTCACGCGCTGAGATCACCCGCATCTCTCTACAAAGTAAAGTCTCTCCGT	120	
QY	1674	aagggcgctgaagaagacaatggaactcttaagaagctctcttgcaaaacaacagatagcgc	1733	
Db	121	AAAGGCGCTGAAGAACATGAGACTCTTAGAAGACTCTTGACAAACACAGTATG6CATC	180	
QY	1734	agcagcaagctaccgagcttcggtgaagcttgtaagcgtgcgcacttaaccaactactgat	1793	
Db	181	ACCAGCAAGTACTCGCGGCTCGGTGTAAGTGTGTACGTCGACCTTACCAACACTGTTAT	240	
QY	1794	agtcaatacttgggaagatactactcctcggaaccccgctcaagagltcaacgltctctt	1853	
Db	241	ACTCAATACTTGGGAAAGTATCACTCCGGAACCCCGCTCAAGAGTTCACCGTCTCTTT	300	
QY	1854	gatactggtctctgcaactctggtcttcctctatactctgaagaagaaatgctgcgaag	1913	
Db	301	GATACTGGTCTCTGACTCTCTGGGTCTCCCTTATCTACTGCAAGACATGCTGCAAG	360	
QY	1914	aaccaccaaaagatctgatactcggaagaagtcgtccacacttcacaaactttaggcaaaccttg	1973	
Db	361	AACCAACCAAAAGTTCGATCCGAGAAAGTGTGTCCACTTCCAGAACTTGAAGCAACCTTG	420	
QY	1974	tctatacaactacgylacagtgtagcatlgaagaagactcttagagctatgatataccgtcagtc	2033	
Db	421	TCTATAACACTAGCGTACAGGTAGTCATGCAAGAAGAACTTAGGCTATGATACGTCAC	480	
QY	2034	tcccaacatttgtagaattccaagaagcagtagactttagaaccacaagaaccggtgtagtc	2093	
Db	481	TCCAACATTGTGACATTCACAGACAGTAGGACTTTACACCAACCAAGTATATGC	540	
QY	2094	tccacctaagcaaatcgaatgcacctgtgtatggcatataccaactgcgtcgcgctcgaag	2153	
Db	541	TTCACTATGCAAGATTCGATGCGATCTCTGTTATGGCATACCCATCCGTCGCGTCAAG	600	
QY	2154	tactcgatacctgtgtttgacaacalgaatgaaccgacacctaagtagtctcaagactgttc	2213	
Db	601	TACTCGATACCGTGTGTTGACAACTGATGAAGCAACGACACTGTGTGCTCAAGACTGTGTC	660	

QY 2214 tcggtttacacgaggaatgagcagagagacatgctcaacgcttgagcatatgatcca 2273
|||||
Db 661 TCGGTTTACATGAGAGGAATGCGAGAGAGCATGCTCAGCGCTTGAGGATATTGATCCA 720
QY 2274 tctactacacagagatctctcaactgggttccagtcactgtgcagcagtaactgcaattc 2333
|||||
Db 721 TCCACTACACAGAGATCTCTTCACCTGGGTTCCAGTCACCTGACACTACTGCGCAATTTC 780
QY 2334 actgtgacagatgcacacaaagcgtgtggtttgttcgatgtgaagggtgagatgtcaact 2393
|||||
Db 781 ACTGTGACAGTGTACCATATGCGGTGTGTGTGTGATGTGAGAGGTGATGTCAAGCT 840
QY 2394 atcttgataccggttagctccaagctgggttcggaactaagcagcagacatctcaacatlcag 2453
|||||
Db 841 ATCTTGATACCGGTAGCTCCAGAGCTGGTCCGACTACAGAGCATTTCTCAACATTCAG 900
QY 2454 caagctatlgagccacacagaaacagtaagtgagtttgacataagattgcgacaactc 2513
|||||
Db 901 CAAGCTATTGGAGGCCACACAGAACCCAGTACGTTGACATAGATTGGGACAACTTT 960
QY 2514 agctacatgctacagatgtgtcttgagatacagcagcagagatgtaaccatgacccctcc 2573
|||||
Db 961 AGCTACATGCTTACAGTTGTCTTTGAGATCAACGGCAGATGTACCCAGTACCCCTTC 1020
QY 2574 gctctacacagcagatcaagggttctgcacacagtgatctcagagtgagaaacatctcc 2633
|||||
Db 1021 GCTATACACAGCAGATCAAGGGTTCGACACAGTGATTCAGAGAGAGAACCATTC 1080
QY 2634 cagaaatgagatcttgagagatgtgtcattcgtgtgactacacagcgtctcttgacagggcc 2693
|||||
Db 1081 CAGAAATGAGCTTGGGAGATGTGTTCATTCGTGAGTACTACAGCGCTTTGACAGGGCC 1140
QY 2694 aacaacctcgttgagcagtaagtaagcaactcga 2726
|||||
Db 1141 AACAACTCTGTTGGGCTAGCTAAAGCAATCTGA 1173

RESULT 4
AX343913
- LOCUS AX343913 1415 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 5 from Patent WO0200899.
ACCESSION AX343913
VERSION AX343913.1 GI:18491959
KEYWORDS
SOURCE Phaseolus vulgaris.
ORGANISM Phaseolus vulgaris.
REFERENCE 1 (sites)
AUTHORS Angenon, G., de Jaeger, G., Goossens, A. and Depicker, A.
TITLE Heterologous gene expression in plants
JOURNAL Patent: WO 0200899-A 5 03-JAN-2002;
FEATURES
source location/Qualifiers
1. 1415
/organism="Phaseolus vulgaris"
BASE COUNT 469 a 259 c 172 g 515 t
ORIGIN

Query Match 26.5% Score 1047; DB 6; Length 1415;
Best Local Similarity 99.6% Pred. No. 0;
Matches 1367; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 8 aatcatgttactcccaatcatcatatagtgaaaatttggtctctcgcggtgtttc 67
|||||
Db 44 AATCATGTACTCCCATATCATATATAGTGAAGTTTGGCTCTCTGCGCGTGTTT 103
QY 68 ttacctctattaaagggtttcccaactaaattctgtatcatctcacttacttg 127

Db 104 TTACCTCATTTTAAAGGGGTTTCCACCTAAATAATTCGATATCATTCCTTACTTTCG 163
QY 128 ttacttaattctcatalaactcttggtgaattatacaccgtctccgacagatactcc 187
Db 164 TTACTTTAATTTCTCATATATCTTGGTGAATTTATTCAGGCTTCGACAGATATTCCT 223
QY 188 acaatttattatttgtttaaacaatttcaaaccgacataaaattatgaagtcgcgtcta 247
Db 224 ACAATTTATATTATTGTTAAACATTTTCAACCCGATTAATTTATGACGCGCTTA 283
QY 248 tcttaatgtgtcctaacatttcaatltgaatataataattacttaattttagcgttg 307
Db 284 TCTTTAATGTAGTCTMAACATTTTCATTTGAATTAATTAATTTACTTTAATTTAGCCGTTG 343
QY 308 gtagaagcataaagatttattcttattcttctctcatalaataatgtttaatacacaata 367
Db 344 GTAGAACCATATAGATTATTTATTTCTTCTTCATTTAAATGTTTAATATACATATTA 403
QY 368 acaaatctcttacctlaagaagatctcccatlttataatttlaaaaaatatattalcaa 427
Db 404 AACAAATTTCTTACTTATAGAAAGATTTCCATTTTATATTTTAAATATATTATTCAA 463
QY 428 ataatttcaaccagtaaatctcaataataatagtggttccaagaatgaataaatttaa 487
Db 464 ATATTTTCAACACAGTAAATCTCATTAATTAATTAATTTGTTTCAAAAAGTAATAATTTAA 523
QY 488 ctccaatttcttattcgcactgacttgaagcaaacccagtgacacacttagcatt 547
Db 524 CTCCTAATTTTATTTATTCAGTGAATCTTAAAGCAACCCAGTGACACATTCACCTT 583
QY 548 ttctcttgataaaaaaaccatltatcattgtatttlttlaacaatgaataatttc 607
Db 584 TTTTCTTTGGATTAATAAATTCATTTATTTTATTTTATACAAAGAAATTTTC 643
QY 608 accaaacaatcaattgtgtgattctctgaagcaagtcagtatgcaaaatctataatc 667
Db 644 ACCAAACATATTTGTGTATTTCTGAAAGCAAGCATATGCAAAATTCATTAATTC 703
QY 668 ccatttgacactgaagagtaacttgaagatctgctcttgaactgcgaagacacatctctcaa 727
Db 704 CCATTGACACTACGGAAGTATCTGAGATCTGCTTTTACATGCGACACACTCTTTCAA 763
QY 728 agtaatttaataatagtaactatataccaagattcatalataccaataactcaattact 787
Db 764 AGTATTTTAAATATAGTTACTATATTTCAGATTTTCATATATTCAAATCTCAATTTACT 823
QY 788 tctaaaaaattaaatagatataataataatattacttctttaaatttaagttaattgtc 847
Db 824 TCTAAAAATTAATTAATGATTAATTAATAATTAATTTACTTTTAAATTTAAGTTAATTTGTT 883
QY 848 gaattgtgacattgatttattcttactatgtttaaattgttlaagatagatttaa 907
Db 884 GAATTTGGACTATTGATTTATTTATTTACTATGTGTTTAAATTTGTTTATAGTACTTTAA 943
QY 908 agtaaatlaaagtaagtagtagagtgtagtagtglttaaccctaaaccaataactaaaca 967
Db 944 AGTAATATATAGTATGATAGTAGAGTGTAGAGTGTACCCCTAAACCATTAACCTATACA 1003
QY 968 ttatgtgtgactaaattcatalataattcttaattgctttaaacttctcttggtatgtaag 1027
Db 1004 TTTATGTGTGACTAATTTTCAATATATTTCTTAATTCCTTTTACCTTTTCTGTATGTAG 1063
QY 1028 tccgtaactagaatagagtggtgacatagcactgtgtgtcttgggttcaatgagt 1087
Db 1064 TCCGTAACTAGATTTACTGTGGGTTCCTATGCGACTCTGTGTGCTTTTGGTTCAATGCTG 1123
QY 1088 ggt-cttgcgcaagaaaaagacaaagaaacaaagaaaaagacaaacagagagacaaac 1146
Db 1124 GATGCTTCGCGAAGAAAGAACAAAGAACAAAGAAAGAAAGAAAGACAAAGAGAGACAAAC 1183
QY 1147 gcaatcacaaaccaactcaaattagtcaactggtcgtgataagatcgcggttcaatgat 1206
|||||

Db 1184 GCATTCACAAACAACCACTCAATTAAGTACGTGGCTGATCAAGATCCGCCGTCCATGTAT 1243
Qy 1207 gctaaatgcataagcaaacagcgccttaacatgacacttaaaatggcaccatct 1266
Db 1244 GCTTAATGCCATGCAACCAACACGCTGCTTAACATGCACTTAAATGGCTCACCCATCT 1303
Qy 1267 caaccacacaaacacacatlgcctttcttctcatcatcacaacacacgtatataat 1326
Db 1304 CAACCCACAGCAAAACACATTCGCTTTTCTTCATCATCACCAACACACCTGTATATAT 1363
Qy 1327 tcatctctccgcgcaccccaattctctcatcacaacacgtcaactgca 1378
Db 1364 TCATTCTCTCCGCACTCAATTTCTTCATCAACACACGTCACACTGCA 1415
RESULT 5
LOCUS AX252300 2970 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 2 from Patent WO0168887.
ACCESSION AX252300
VERSION AX252300.1 GI:15985641
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 2970)
AUTHORS Jung, R. and Kinney, A.J.
TITLE Hypoallergenic transgenic soybeans
JOURNAL Patent: WO 0168887-A 2 20-SEP-2001;
E. I. du Pont de Nemours and Company (US); PIONEER HI-BRED
INTERNATIONAL, INC. (US)
FEATURES
source 1..2970
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Chimeric construct"
BASE COUNT 1018 a 547 c 527 g 878 t
ORIGIN
Query Match 22.4%; Score 887; DB 6; Length 2970;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 887; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2725 aagttgaactaaataagctgaagtgtaagagctcatgagagagctgtaaatgtatc 2794
Db 1796 AAGTATGAACATAAATGCAATGTAGGTGAAGAGCTCATGAGAGCATGGAATATGTATTC 1855
Qy 2795 cgacacatgaacagataaataactagctcactcactcactctctcactgataaacaag 2854
Db 1856 CGACCATGTAAACAGTATATTAACAGTCCATCTCATTCTTCTATGATTAACAAAG 1915
Qy 2855 atgtatgatataataaacctctatctatgacacctattgtctctatgataaattcctc 2914
Db 1916 AGTATATGATATATTAACACTATCTATGACACTTATGTTCTATGATTAATTTCTCT 1975
Qy 2915 tatctttaaataactcgaacgcgagagcgtatgagaaatgcttcaaatagtaaaaaaac 2974
Db 1976 TATTTTATTAATCATCTGAATCGTAGCGCTTATGGAATGCTTCAATATGTAACAAAAC 2035
Qy 2975 aaatgttactaataagacttctaacaacatctaacttagactatgtgaacgagacata 3034
Db 2036 AAATGTGATCTTAAGACTTCTTAACATCTTAACACTTGAACGAGACATTA 2095
Qy 3035 gtgttaagaagacataaataataatggaagaagtgtgtctccatctatataatata 3094
Db 2096 GTTTAAGAGACATTAACATTAATGGAAGAAGTTGTCTCCATTTATATATATATA 2155
Qy 3095 ttaccacactatgtatataatagagtggttaagagagacataaacaattataagaagaag 3154
Db 2156 TTACCACACTTATGATATATTAATGAGATGTTAAGAGACATTAACATTTAAGAGAGAG 2215
Qy 3155 ttgtatcatctatataatataactaccacttataatataatataccactatct 3214

Db 2216 TTTGTATTCATTTATATATATATATACCATTTATATATATATATATTCACCTATTT 2275
Qy 3215 aatgtcttataaagtttgatcacaatgatattcttaaatatgaatgtatgatata 3274
Db 2276 AATGTCTTATATAGGTTTATATCATGATATTTCTATATATTTAGTTGATATATATGAA 2335
Qy 3275 aggtactactatgaactcctcactcctcgtatgaaggttgatcacttaagttggtc 3334
Db 2336 AGGTTACTATTTGAACTCTTACTGTATTAAGGTGATCATCTTAAGTGGCTC 2395
Qy 3335 atttaatttatatgtctctctacagataaaaaaaatlaaggttggttgatataaata 3394
Db 2396 ATTTATATTTATATGCTTCTTACAGATTAATAAAAAAATATGATGCTTGAATAAATAT 2455
Qy 3395 tgaaggaatttaaaataataataataataacataacataatcgtatataaattat 3454
Db 2456 TGAAGGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2515
Qy 3455 tataataacatttactataaaaaaagaataatgctataaactatatacactgttca 3514
Db 2516 TATATATTAACATTTATATGTAATAAAGTAATATTTGCTATTAATCTATACATTCGTTA 2575
Qy 3515 gccttgctgacgactcactaatatttaacgagagtaaacatatgtgacttttgta 3574
Db 2576 GCCTTGCTGAGACACTCTCAATTTATTAACGAGACTTAACATATTTGACTTTTGCTTA 2635
Qy 3575 tttaacaattatatttaacactatagaatatttttttttalc 3621
Db 2636 TTTAACAAATTTATTTATTAACACTATATGAATTTTATTTTATATC 2682

RESULT 6
LOCUS E00452 1990 bp DNA linear PAT 29-SEP-1997
DEFINITION DNA sequence coding for faseorin.
ACCESSION E00452
VERSION E00452.1 GI:2168735
KEYWORDS JP 1985210988-A/2.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
REFERENCE 1 (bases 1 to 1990)
AUTHORS Ujyon,D.K., Yelmoshii,S.H., Jierli,E.S., Denisu,D.S. and
NORMOTO,M.
TITLE DEVELOPMENT OF PLANT STRUCTURAL GENE
JOURNAL Patent: JP 1985210988-A 2 23-OCT-1985;
AGURIJENETIKUSU RES ASSOC LTD
COMMENT
OS Escherichia coli
PN JP 1985210988-A/2
PD 23-OCT-1985
PF 16-APR-1984 JP 1984077452
PR 15-APR-1983 US 83 485614
PI JIJON DEJI KENPU, TEIMOSHII SHII HOORU, JIERII ERU SURAITOMU,
PI DENISU DABURILYU SATSUNON, NORIMOTO MURAI
PC C12N15/00,A01H1/00,C12N1/20,C12N5/00,(C12N1/20,C12R1:01), PC
(C12N5/00,
PC C12R1:91);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC "source": library=plasmid PK54-KB;
FH key Location/Qualifiers
FT 5'UTR 1..177
FT CDS
FT join(178..490,563..753,842..922,1047..1277,1406..1664, FT
1768..1958)
FT /product="faseorin"
FT promoter 101..177

FT exon 178..490
FT intron 491..562
FT exon /note='IVS1,72bp'
FT exon 563..753
FT intron 754..841
FT exon /note='IVS2,88bp'
FT intron 842..922
FT intron 923..1046
FT exon /note='IVS3,124bp'
FT intron 1047..1277
FT intron 1278..1405
FT exon /note='IVS4,128bp'
FT exon 1406..1664
FT intron 1665..1767
FT exon /note='IVS5,103bp'
FT exon 1768..1958
FT 3'UTR 1958..<2090.
FEATURES
Source Location/Qualifiers
1..1990
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 614 a 416 c 372 g 588 t
ORIGIN

Query Match 4.6%; Score 181; DB 6; Length 1990;
Best Local Similarity 100.0%; Pred. No. 9.9e-77;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1377 catatgctgtatccatgcccaatctccatgcatgttccaaaccctctctctat 1436
DB 1 CATATGCCGTATCCATGCCCAATCTCCATGCAATGTTCCACCACTTCTCTTAT 60
OY 1437 ataataaccataataaccataatacaactcctcttcacatccatccacagat 1496
DB 61 ATAAATACCTATAATACCTATAATACACTCTTTCATCATCCATCCATCCAGAGT 120
OY 1497 actactactactactactataatacccccaaccacatcatatcaatactactatg 1556
DB 121 ACTACTACTCTACTACTATAATATATATCCCAACCACTCATATTCATATCTACTATG 180
OY 1557 a 1557
DB 181 A 181

RESULT 7
LOCUS A06496 2090 bp DNA linear PAT 10-NOV-1993
DEFINITION Artificial sequence for phaseolin.
ACCESSION A06496
VERSION A06496.1 GI:490411
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 2090)
AUTHORS Kemp,J.D., Hall,T.C., Slightom,J.L., Sutton,D. and Murali,N.
TITLE Plant structural gene expression
JOURNAL Patent: EP 0126546-A 24 28-NOV-1984;
LUBRIZOL GENETICS INC
FEATURES
Source Location/Qualifiers
1..2090
/organism="synthetic construct"
/db_xref="taxon:32630"
join(178..490,563..753,842..922,1047..1277,1406..1664,
1768..1958)
/codon_start=1
/transl_table=11
/product="phaseolin"
/protein_id="CAA00562.1"
/db_xref="GI:490412"
/translation="MMARAVPLLILGILFLASLSASPTSLREESQDNPPYFNSDN

SMNTFEKNOYGHTRVLORPDOOSKRLONLSDRYLVERSKRPETLLLPQADAEILLV
RSSGAILVIVKPPDDRREFFLTSDNPJESPDHOKTPACTIFYLVNPDKEKRLITQLAM
PVNPPQIHEFFFLSTEAQOSTLOEFSKHILLPASNKSPFEELINRYLFEEBOQOEVIYN
IDSEQIKELSKHAKSSSRKSLSKQDNITGNFGLTERTDNLNLVLSITMEGALF
VPHYYSKAIIVLVNCEGAHEVLVGPNGKETLEFEYSRELSDDFVIPAAPVAI
KATSNVNETGFGIANNNNRNRLLAGKTDNYISIGRALDKVDVILGTFSGSGEVMNL
INROSSYFVDAAHHHQQEQKGRGARVY"
<178..490
/number=1
intron 491..562
/number=1
exon 563..753
/number=2
intron 754..841
/number=2
exon 842..922
/number=3
intron 923..1046
/number=3
exon 1047..1277
/number=4
intron 1278..1405
/number=4
exon 1406..1664
/number=5
intron 1665..1767
/number=5
exon 1768..>1958
/number=6

BASE COUNT 645 a 423 c 397 g 625 t
ORIGIN

Query Match 4.6%; Score 181; DB 6; Length 2090;
Best Local Similarity 100.0%; Pred. No. 9.8e-77;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1377 catatgctgtatccatgcccaatctccatgcatgttccaaaccctctctctat 1436
DB 1 CATATGCCGTATCCATGCCCAATCTCCATGCAATGTTCCACCACTTCTCTTAT 60
OY 1437 ataataaccataataaccataatacaactcctcttcacatccatccacagat 1496
DB 61 ATAAATACCTATAATACCTATAATACACTCTTTCATCATCCATCCATCCAGAGT 120
OY 1497 actactactactactactataatacccccaaccacatcatatcaatactactatg 1556
DB 121 ACTACTACTCTACTACTATAATATATATCCCAACCACTCATATTCATATCTACTATG 180
OY 1557 a 1557
DB 181 A 181

RESULT 8
LOCUS PVAPBASE 4764 bp DNA linear PLN 07-APR-1993
DEFINITION Phaseolus vulgaris gene for alpha-phaseolin.
ACCESSION X52626
VERSION X52626.1 GI:20972
KEYWORDS alpha-phaseolin; glycoprotein; phaseolin; seed storage protein.
SOURCE Phaseolus vulgaris.
ORGANISM Phaseolus vulgaris.
REFERENCE 1 (bases 1 to 4764)
AUTHORS Anthony,J.L.
TITLE Direct Submission
JOURNAL Submitted (17-APR-1990) Antony J.L., Dept. of Biology, Texas A & M
University, College Station, TX 77843-3258, USA
REFERENCE 2 (bases 1987 to 4764)

ORGANISM Phaseolus vulgaris
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
1 (bases 1 to 1475)
AUTHORS Slightom,J.L., Drony,R.F., Klassy,R.C. and Hoffman,L.M.
TITLE Nucleotide sequences from phaseolin cDNA clones: the major storage
proteins from Phaseolus vulgaris are encoded by two unique gene
families
JOURNAL Nucleic Acids Res. 13 (18), 6483-6498 (1985)
MEDLINE 86041851
COMMENT The single base deletion in a variant clone can be due either to a
cloning artefact or to a represented pseudogene.
FEATURES
source 1. 1475
/organism="Phaseolus vulgaris"
/db_xref="taxon:3885"
misc_RNA 1
/note="cap site"
1. 1475
/note="mRNA"
4
misc_RNA /note="cap site in variant clone"
78. 1343
/note="precursor"
/codon_start=1
/protein_id="CAA26789.1"
/db_xref="GI:21040"
/db_xref="SWISS-PROT:P02853"
/translation="MWRKRPPLLGLFLFLASLSASRATSLREEEGODNPFYNSDN
SWNTLFKNQGHIVLQRFDOQSRKLQNLDEYRLVERRSKPEITLLPQDAEILLV
RSGSAIIVLYKPDREYFELTDNPIFSDBQKIPAGTIFLYVDPKEDRIILQIAM
VGNPQIHEFLFSTEAQSYLQEFKSHLEASNSKFEELNRLFLFEEGGQEEIVN
IDSDQIKELSHAKSSRSKSLSKODNTGNFGLNTERDNLVNLVLSIMEGALF
VPHYSKRAIVLYNCEGAHYEIVGPKNGKTELEESYRATLSKDDVPIPAAYPAI
KAISVNFTEGCIYANNNNRNLAKGTDNVTSSIGRALDGDVGLITFSSGDEVML
INKSGSYFDVAHHHQEQKGRKGAIFY"
78. 146
/note="putative"
147. 1340
/product="mature beta-phaseolin (aa 1-398)"
misc_feature 149. 150
/note="pot. altern. signal peptide cleavage site"
misc_feature 152. 153
/note="pot. altern. signal peptide cleavage site"
misc_feature 155. 156
/note="pot. altern. signal peptide cleavage site"
variation 170
/note="A is missing in variant clone causing frameshift
and premature stop codon"
831. 839
misc_feature /note="pot. glycosylation site"
1098. 1106
misc_feature /note="pot. glycosylation site"
1454. 1458
misc_feature /note="put. polyadenylation signal"
1475
/note="polyadenylation site"
BASE COUNT 455 a 335 c 306 g 379 t
ORIGIN

Query Match 3.4%; Score 133; DB 8; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.6e-53;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2732 aataagatgaactaaatgcatgtagtgcgaagcgaagcgaagcgaatgct 2791
|||||
Db 1343 AATAAGTATGAACATAAATGCAATGTAGTGAAGACCTCATGAGAGCAATATTCG 1402
|||||
Qy 2792 atccgacatgttaacagctataactgagctccatctcattcttatgataaaca 2851
|||||

Db 1403 ATCCGACCATGTACAGATATATACTGAGCTCCATCTCACTTCTTCATGATATAACAA 1462
Qy 2852 agatgttatgat 2864
|||||
Db 1463 AGGATGTTATGAT 1475

RESULT 11
ZMZEINI 207 bp DNA linear PLN 30-MAR-1995
LOCUS Zea mays/P.vulgaris chimeric beta-phaseolin/zein gene 5' end region.
DEFINITION X06175
ACCESSION X06175.1 GI:22546
VERSION X06175.1 GI:22546
KEYWORDS phaseolin; signal peptide; zein protein.
SOURCE Zea mays.
Zea mays.
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 207)
AUTHORS Hoffman,L.M., Donaldson,D.D., Bookland,R., Rashka,K. and
Herman,E.M.
TITLE Synthesis and protein body deposition of maize 15-kd zein in
transgenic tobacco seeds
JOURNAL EMBO J. 6, 3213-3221 (1987)
COMMENT This sequence is a chimera between Phaseolus vulgaris sequences
from J.L. Slightom, S.M. Sun and T.C.Hall (1983) 'Complete
nucleotide sequence of a French bean storage protein gene:
phaseolin', Proc. Natl. Acad. Sci USA 80:1897-1901, and Zea mays
sequences from
K. Pedersen, P. Argos, S.V.L. Naravana and B.A. Larkins (1986)
'Sequence analysis and characterization of a maize gene encoding a
high-sulfur zein protein of Mr 15,000',
J. Biol. Chem. 261:6279-6284.
FEATURES
source 1. 207
/organism="Zea mays"
/db_xref="taxon:4577"
misc_feature 1. 95
/note="phaseolin sequence"
9. 14
/note="TATA element"
41
misc_feature /note="phaseolin transcription initiation site"
77. 81
/note="TATA element"
96. >207
misc_feature /note="zein sequence"
98
/note="zein transcription initiation site"
161. >207
CDS 161. >207
/codon_start=1
/product="chimeric beta-phaseolin/zein"
/protein_id="CAA29542.1"
/db_xref="GI:22547"
/translation="MKWIVLVVCAISAA"
BASE COUNT 61 a 60 c 33 g 53 t
ORIGIN

Query Match 2.4%; Score 95; DB 8; Length 207;
Best Local Similarity 100.0%; Pred. No. 1.1e-34;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1437 aataaccataataaccataataaccataaccataccatcttcacatccatccacagagt 1496
|||||
Db 1 ATATATCCTATTAATACCTCTAATATACCTCACTTCTTCATCATCATCCATCCAGACT 60
|||||
Qy 1497 actactactactactactataataaccacccaac 1531
|||||
Db 61 ACTACTACTCTACTACTATATATACCCCAACCCMAC 95
|||||

RESULT 12
PVPHASAR
LOCUS PVPHASAR 1535 bp mRNA linear PLN 21-MAR-1995
DEFINITION Phaseolus vulgaris mRNA for alpha-type phaseolin.
ACCESSION X02980
VERSION X02980.1 GI:21035
KEYWORDS direct repeat; glycoprotein; phaseolin; seed storage protein;
signal peptide; storage protein; tandem repeat.
SOURCE Phaseolus vulgaris.
ORGANISM Phaseolus vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
REFERENCE 1 (bases 1 to 1535)
AUTHORS Slightom,J.L., Drong,R.F., Klassy,R.C. and Hoffman,L.M.
TITLE Nucleotide sequences from phaseolin cDNA clones: the major storage
proteins from Phaseolus vulgaris are encoded by two unique gene
families
JOURNAL Nucleic Acids Res. 13 (18), 6483-6498 (1985)
MEDLINE 86041851
COMMENT The variability in transcription initiation may be due to multiple
and overlapping TATA-elements in the genomic sequence (see
<PVPHASL2>). Seven slightly different alpha-phaseolin cDNA clones
were detected. Whether the differences represent nucleotide
substitutions in allelic genes or show divergences of non-allelic
genes is unknown.

FEATURES
source location/Qualifiers
1..1535 /organism="Phaseolus vulgaris"
/db_xref="taxon:3885"
1 /note="cap site"
misc_RNA 10
/note="cap site in variant clones"
misc_RNA 13
/note="cap site in variant clone"
CDS 87..1397
/note="precursor"
/codon_start=1
/protein_id="CAA26718.1"
/db_xref="GI:21036"
/db_xref="SWISS-PROT:P07219"
/translation="MMARVPLLLGLTFLASLSASPATSLRFEESODNPFYNSN
SMNTLEKNQGHIFVILQRPDOQSKRLNLEDRYRERSKRETLILPQDAELLIV
RSGSAILVLPKPPDRKREFFLTQGNFTSDNQTIPAGTITLVNPPKEDLRIQLA
MPVNPQIHFEFLSTEAQOSYLOEFKRLLEAFNSKFEELNVLEEEQOEQGO
EGVIVNDSQIELSTKHAQSSSRKSHKODNTIGNFNGTLRTDNLNLLSIEH
KEGALFVPHYSKAILVILVNEGFAHVELVGPKNKFTLEFESYRAELSKDQVPIA
AYPVAKATSNVNTFGINANNNNRNLNLAGKTNYVLSIGRALDGDVGLTFSGSG
EYWKILNKSGSFYVDGHHHQDQDQSGHQEQKRGKAFV"
sig_peptide 87..1155
/note="putative"
mat_peptide 156..1394
/product="put. mature alpha-phaseolin (aa 1-413)"
misc_feature 158..159
/note="pot. altern. signal peptide cleavage site"
misc_feature 161..162
/note="pot. altern. signal peptide cleavage site"
misc_feature 164..165
/note="pot. altern. signal peptide cleavage site"
variation 339
/note="C is T in variant clones"
variation 456..464
/note="ACGCAAGC (Thr, Gln, Gly) is ACCAGC (Thr, Ser) in
variant clones"
variation 456..464
/note="ACGCAAGC (Thr, Gln, Gly) is GCTAGC (Ala, Ser) or
ACTAGC (Thr, Ser) in variant clones"
variation 486..488
/note="AAC (Asn) is CAC (His) in variant clones"
variation 533

variation /note="C is T in variant clones"
594..596
/note="GAA (Glu) is GAC (Asp) in variant clones"
repeat_region 711..725
/note="direct repeat 1"
repeat_region 726..740
/note="direct repeat 1"
misc_feature 858..866
/note="pot. glycosylation site"
misc_feature 1125..1133
/note="pot. glycosylation site"
repeat_region 1321..1348
/note="direct repeat 2"
repeat_region 1349..1375
/note="direct repeat 2"
misc_feature 1509..1514
/note="put. polyadenylation signal"
misc_feature 1530
/note="polyadenylation sites in variant clones"
BASE COUNT 480 a 344 c 323 g 388 t
ORIGIN

Query Match 2.3%; Score 90; DB 8; Length 1535;
Best Local Similarity 100.0%; Pred. No. 2,4e-32;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1468 actcttcacatccatccatccagagtactactactactactataataccccaac 1527
|||||
DB 1 ACTCTTTCATCATCCATCCATCCAGAGTACTACTACTACTATAATACCCCAACC 60
OY 1528 caactcatatcaactactactactatga 1557
|||||
DB 61 CAACTCATTTCAATCACTACTACTACTATGA 90

RESULT 13
LOCUS U01131 1454 bp mRNA linear PLN 16-MAY-1996
DEFINITION Phaseolus vulgaris Sanilac clone 1-12 phaseolin (Phs) mRNA,
complete cds.
ACCESSION U01131
VERSION U01131.1 GI:403593
KEYWORDS
SOURCE
ORGANISM French bean.
Phaseolus vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
REFERENCE 1 (bases 1 to 1454)
AUTHORS Kaml,T.A. and Gepts,P.
TITLE Phaseolin nucleotide sequence diversity in Phaseolus. I.
Intraspecific diversity in Phaseolus vulgaris
JOURNAL Genome 37, 751-757 (1994)
MEDLINE 95095072
REFERENCE 2 (bases 1 to 1454)
AUTHORS Kaml,T.A.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1993) James A. Kaml, Agronomy and Range Science,
University of California at Davis, Davis, CA 95616, USA
FEATURES
source location/Qualifiers
1..1454 /organism="Phaseolus vulgaris"
/cultivar="Sanilac"
/db_xref="taxon:3885"
/map="linkage group D7"
/clone="Sanilac 1-12"
/haplotype="S" type beta-phaseolin"
/tissue_type="cotyledon"
/clone_lib="Sanilac PUC19 cDNA"
/dev_stage="cotyledonary stage"
1..1454

gene

```

5'UTR      /gene="Pns"
           1..53
           /gene="Pns"
           54..1319
           /gene="Pns"
           /function="seed storage protein"
           /codon_start=1
           /product="phaseolin"
           /db_xref="GI:403594"
           /db_xref="AA09534.1"

CDS
           /translation="MVARAVPLMLGLTFLASLSASFATSLREEESODNPEFNSDN
           SNNLTPEKNOYGHTRVROPRDOOSKLONIEDRVLVFRSKPETLLIPQADALLLV
           KSGSALIVLKPDDRREYFFLSDNPFSDHQTIPAGTIFYLYNPPKEDIRIOLAM
           PVNNPQIHDFLSTEAQSYLOEFKSHLEASFNSKFETINVLFEBSQDEGVIN
           IDSEQIKELSKHAKSSRSKSLSKQDNTIGNFENGLNEDNSLNLVLEEMKALF
           VPHYYSKAIYILVNEGGAHVELVGPKNKTELEYSYRAELSKDVFVIPAAPVAI
           KATSNVPTFGTGINANNNNNLGAKTDNYSISGRALDKDVLGLTFSGSGDEVNKL
           MKQSSYFVDHHDHQQEOKGKKAFFV"
           54..125
           /gene="Pns"
           123..1316
           /gene="Pns"
           /function="seed storage protein"
           /product="phaseolin"
           674..679
           /gene="Pns"
           /standard_name="T' type alpha-phaseolin"
           /label="15bp-repeat"
           /phenotype="T' type phaseolin"
           /replace="gaggaaggagacgaagaaggagagacagca"
           /product="phaseolin"
           807..815
           /gene="Pns"
           /function="putative glycosylation site"
           1074..1082
           /gene="Pns"
           /function="putative glycosylation site"
           1271..1298
           /gene="Pns"
           /standard_name="T' type alpha-phaseolin"
           /label="27bp-repeat"
           /phenotype="T' type phaseolin"
           /replace="tcaccaacaggaacgaagaagggaatcaccacaaggacgaagaag
           gga"
           /product="phaseolin"
           1320..1454
           /gene="Pns"
           1429..1434
           /gene="Pns"

3'UTR
polyA_signal

BASE COUNT      451 a      306 g      373 t
ORIGIN
Query Match      1.9%: Score 77; DB 8; Length 1454;
Best Local Similarity 100.0%; Pred. No. 5.2e-26;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2789 tgtatccagcatgtaacataactgagccatctcaactcttctatagaataa 2848
|||||
Db 1375 tctatccagcatgtaacataactgagccatctcaactcttctatagaataa 1434
|||||

OY 2849 caaagatgatatgata 2865
|||||
Db 1435 CAAAGATGTTATGATA 1451
|||||

RESULT 14
ZMEIN2      348 bp      DNA      linear      PLN 06-AUG-1992
LOCUS      Maize chimeric zein/beta-phaseolin gene 3'end region.
DEFINITION
ACCESSION  X06176
VERSION     X06176.1 GI:22548
KEYWORDS    phaseolin; zein protein.

```

```

SOURCE      Zea mays.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Eukaryota; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoidae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 348)
AUTHORS    Hoffman,L.M., Donaldson,D.D., Bookland,R., Rashka,K. and
            Herman,E.M.
TITLE       Synthesis and protein body deposition of maize 15-kd zein in
            transgenic tobacco seeds
JOURNAL     EMBO J. 6, 3213-3221 (1987)
FEATURES    Location/Qualifiers
            source          1..348
            misc_feature    /organism="Zea mays"
            misc_feature    /db_xref="taxon:4577"
            misc_feature    <1..269
            misc_feature    /note="zein sequence"
            misc_feature    192..197
            misc_feature    /note="polyA signal"
            misc_feature    236..241
            misc_feature    /note="polyA signal"
            misc_feature    253..258
            misc_feature    /note="polyA signal"
            misc_feature    270..348
            misc_feature    /note="phaseolin sequence"
            misc_feature    298..303
            misc_feature    /note="polyA signal"
            polyA_site       338
            polyA_site       /note="pot.polyA site"
            polyA_site       340
            polyA_site       /note="pot.polyA site"

BASE COUNT      93 a      93 c      62 g      100 t
ORIGIN
Query Match      1.9%: Score 75; DB 8; Length 348;
Best Local Similarity 100.0%; Pred. No. 5.9e-25;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2819 gagctcatctcaactcttctcatgaataaagaagatgatatataatacacta 2878
|||||
Db 274 GAGCTCATCTCAGCTTCTCTATGATTAACACACTCTA 333
|||||

OY 2879 tctatgacacctat 2893
|||||
Db 334 TCTATGACACCTTAT 348
|||||

RESULT 15
LOCUS      U01121      1478 bp      mRNA      linear      PLN 16-MAY-1996
DEFINITION Phaseolus lunatus phaseolin (Pns) mRNA, complete cds.
ACCESSION  U01121
VERSION     U01121.1 GI:403581
KEYWORDS
SOURCE      Lima bean.
ORGANISM    Phaseolus lunatus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1 (bases 1 to 1478)
AUTHORS    Kaml,U.A. and Gepts,P.
TITLE       Phaseolin nucleotide sequence diversity in Phaseolus. I.
            Intraspecific diversity in Phaseolus vulgaris
JOURNAL     Genome 37, 751-757 (1994)
MEDLINE     95095072
REFERENCE   2 (bases 1 to 1478)
AUTHORS    Kaml,U.A.
TITLE       Direct Submission
JOURNAL     Submitted (30-AUG-1993) James A. Kaml, Agronomy and Range Science,
            University of California at Davis, Davis, CA 95616, USA
FEATURES    Location/Qualifiers

```



```
source      1. .1478
            /organism="Phaseolus lunatus"
            /strain="G25804"
            /db_xref="taxon:3884"
            /clone="G25-39"
            /tissue_type="cotyledon"
            /clone_lib="G25 pUC19 cDNA"
            /dev_stage="cotyledonary stage"
            1. .1478
            /gene="Phs"
            1. .60
            /gene="Phs"
            61. .1347
            /gene="Phs"
            /function="seed storage protein"
            /codon_start=1
            /product="Phaseolin"
            /protein_id="AA099532.1"
            /db_xref="GI:403582"
            /translation="MMRAVPDLLLGLFLASLSPALISLREHNESQNDPFFSSD
            SMOTLFKNQGHIFVLOSFDQSERLQNLDEYRLEVEPMSKPEITLLPOADAPELIVY
            RSSGALLALVPGSTIYSLKQDPTIKIPACTITPFLINPQNNEDIRITKLMTYNNPQ
            IODFLSSTEAOQSYLYGFRKDIIDASFNSPIEETINRLFAEGRQEGVIVNIGSDLI
            QELSRHAKSSSRKSLDHSNLDISNEMGNTDLYVNSLDVLTLYVEIKEGGLFVPHYS
            KAIIVLVEEGVAKVELVGPKREKSELEETRADVSEGDVFIIPAAYPAVKAISNV
            NTFSEGINANNRYRILLTGKGPTGKEDNIIISAGINPDVGLMPGSGEDVQKLFNNQ
            NLSHFVNGSYHKNNQPPHQEQOKQKGRKAFVY"
            61. .132
            /gene="Phs"
            133. .1344
            /gene="Phs"
            /function="seed storage protein"
            /product="phaseolin"
            151. .159
            /gene="Phs"
            /function="putative glycosylation site"
            796. .804
            /gene="Phs"
            /function="putative glycosylation site"
            1063. 1071
            /gene="Phs"
            /function="putative glycosylation site"
            1237. .1245
            /gene="Phs"
            /function="putative glycosylation site"
            1255. 1263
            /function="putative glycosylation site"
            /gene="Phs"
            /function="putative glycosylation site"
            1348. .1478
            /gene="Phs"
            1457. .1462
            /gene="Phs"
            /polyA_signal
            BASE COUNT  441 a  340 c  311 g  386 t
            ORIGIN

Query Match      1.6%: Score 63; DB 8; Length 1478;
Best Local Similarity 100.0%: Pred. NO. 3.4e-19;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1495 gtactactactactataataaccccaaccacatatacatatactactacta 1554
      |||
Db   2  gTACTACTACTACTATAATACCCACCAACCATATTCATATCTACTACTACTA 61
QY  1555 tga 1557
      |||
Db   62  TGA 64
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2002, 01:57:49 ; Search time 791.3 Seconds
(without alignments)
8585.657 Million cell updates/sec

Title: US-09-643-755B-3
Perfect score: 3957
Sequence: 1 ctgcagatcatctgtact.....acctcaactaaggtacc 3957

Scoring table: OLIGO-MUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3957	100.0	3957	22	AA500570
2	1544	39.0	1558	20	AA110376
3	1439	36.4	1547	20	AA950961
4	1226	31.0	1244	20	AA410392
5	1173	29.6	1173	22	AA500568
6	887	22.4	2970	22	AA017528
7	49	1.2	1098	4	AA030063
8	49	1.2	1098	11	AA004683
9	49	1.2	1175	5	AA040295

10	49	1.2	1175	13	AA020949
11	49	1.2	1210	12	AA014051
12	49	1.2	1240	16	AA030006
13	49	1.2	1278	5	AA040055
14	49	1.2	1289	4	AA030022
15	49	1.2	1291	10	AA091157
16	49	1.2	1314	4	AA030049
17	49	1.2	1460	3	AA020043
18	49	1.2	1460	5	AA040180
19	49	1.2	2727	10	AA091188
20	49	1.2	2733	20	AA0206463
21	49	1.2	2982	10	AA091185
22	38	1.0	1290	4	AA030209
23	35	0.9	637	14	AA049459
24	34	0.9	1096	20	AA083966
25	33	0.8	179	5	AA040296
26	33	0.8	186	13	AA020950
27	33	0.8	819	20	AA083967
28	32	0.8	107	12	AA014777
29	29	0.7	1460	3	AA020043
30	29	0.7	1460	5	AA040180
31	25	0.6	33	21	AA062421
32	24	0.6	25	20	AA0210389
33	23	0.6	1083	21	AA097159
34	23	0.6	1140	24	AA044658
35	23	0.6	1814	21	AA037875
36	23	0.6	44242	23	AB019930
37	22	0.6	262	22	AB07288
38	22	0.6	262	22	AA089933
39	22	0.6	3607	18	AA089990
40	22	0.6	5376	24	AB044150
41	22	0.6	6385	22	AA045352
42	22	0.6	14708	22	AA045513
43	22	0.6	14708	22	AA046773
44	22	0.6	14708	24	AB034244
45	22	0.6	15881	24	AB032260

ALIGNMENTS

RESULT 1
AA500570 standard; DNA; 3957 BP.

AC AA500570;
DT 14-MAY-2001 (first entry)

DE Bovine phaseolin promoter pre-pro-chymosin-phaseolin terminator.

XX Chymosin; transcription regulator; terminator sequence; soybean; corn;
KW pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;
KW barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;
KW safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;
KM squash; jojoba; ds; phaseolin; promoter; terminator; mutant; French bean.

XX Chimeric - Bos sp.

OS Chimeric - Phaseolus vulgaris.

XX Key Location/Qualifiers
FH promoter 1..1553
FT /*tag= a
FT /*note= "Phaseolin promoter"

FT CDS 1554..2726
FT /*tag= b
FT /*product= "Bovine pre-pro-chymosin"

FT terminator 2727..3957
FT /*tag= c
FT /*note= "Phaseolin terminator"

XX WO200114571-A1.

Prochymosin (prore
Remin gene. Synt
Chymosin open read
Sequence of prochy
Sequence of veal c
Cloned sequence of
cDNA sequence corr
Pre-proremnin-A ge
Sequence of recomb
BamHI/Sall insert
2.7 Kbp HindIII fr
BamHI insert from
Sequence of p1epro
Prochymosin gene 5
DNA encoding the f
Sequence coding fo
DNA coding for a r
DNA encoding His-c
Remin - casein co
Pre-proremnin-A ge
Sequence of recomb
Phaseolin gene pro
PCR primer used to
Arabidopsis thalia
Human aspartyl pro
Arabidopsis thalia
Drosophila melanog
Human pancreatic c
Human digestive sy
Cryptosporidium pa
Human immune syste
Chemically pretrea
Tumour suppressor
Human immune syste
Human immune syste

Query Match	Best Local Similarity	100.0%;	Score 3957;	DB 22;	Length 3957;
Matches 3957;	Conservative	0;	Mismatches	0;	Indels
					Gaps
01-MAR-2001.					
23-AUG-2000;	2000WO-CA00975.				
23-AUG-1999;	99US-0378696.				
(SEMB-)	SEMBIOSIS GENETICS INC.				
Van Rooijen G,	Keon RG, Boothe J, Shen Y;				
MP1: 2001-226621/23.					
P-PSDB: AAU00536.					
Producing chymosin in seeds of plants such as rice, flax, rape seed, by transforming plant cell with a nucleic acid encoding chymosin operably linked to transcription regulator and terminator sequences -					
Example 1; Fig 2; 56pp; English.					
The sequence represents a chimeric polynucleotide comprising a pre-pro-chymosin, a phaseolin promoter and a phaseolin terminator. Chymosin can be produced in a plant seed through introduction of a chimeric nucleic acid molecule, comprising a nucleic acid sequence encoding a chymosin polypeptide operatively linked to transcription regulator and terminator sequences, into a plant cell. The sequences are useful for producing plant seeds, in particular seeds of soybean, rape seed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, Arabidopsis thaliana, potato, flax/linseed, safflower, oil palm, groundnut, Brazil nut, coconut, castor, coriander, squash, jojoba and rice.					
Sequence 3957 BP, 1263 A; 790 C; 609 G; 1295 T; 0 other;					
1	1	100.0%;	Score 3957;	DB 22;	Length 3957;
1	1	100.0%;	Pred. No. 0;		
1	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels
					Gaps
01	1	100.0%;	Score 3957;	DB 22;	Length 3957;
01	1	100.0%;	Pred. No. 0;		
01	1	Matches 3957;	Conservative	0;	Mismatches
				0;	Indels

QY	541	agccattttttctttggaataaaaaaacccaatatcatctgtattttttatataatga	600
DB	541	agccattttttctttggaataaaaaaacccaatatcatctgtattttttatataatga	600
QY	601	aaattcaaccaaacatcttcttggtatcttcggaagcagcatggtatgcaaatctc	660
DB	601	aaattcaaccaaacatcttcttggtatcttcggaagcagcatggtatgcaaatctc	660
QY	661	ataattccatttgcacatcgcggagtaactggaatcgtctgtcttatactgcgagacat	720
DB	661	ataattccatttgcacatcgcggagtaactggaatcgtctgtcttatactgcgagacat	720
QY	721	cttctaagaatatttaataatagttacatataatccaagttccatataatcaatccaa	780
DB	721	cttctaagaatatttaataatagttacatataatccaagttccatataatcaatccaa	780
QY	781	tattactcttaaaaaatlaattagatataatnaaaatattacttttttaatttaagt	840
DB	781	tattactcttaaaaaatlaattagatataatnaaaatattacttttttaatttaagt	840
QY	841	aattgttgaatttgcacttgaatttatatattctactatglttaattgtttatagat	900
DB	841	aattgttgaatttgcacttgaatttatatattctactatglttaattgtttatagat	900
QY	901	agtttaagaatgaataagaatgaatgagtgagtggtgtttacccataaacataaac	960
DB	901	agtttaagaatgaataagaatgaatgagtgagtggtgtttacccataaacataaac	960
QY	961	tataacattttagtgcgaactaatttccatatacttcttatgctttaccccttcttgt	1020
DB	961	tataacattttagtgcgaactaatttccatatacttcttatgctttaccccttcttgt	1020
QY	1021	atgtaagtcggttaactgaattacagttggttgcacatgcctcgtgtcttttgttc	1080
DB	1021	atgtaagtcggttaactgaattacagttggttgcacatgcctcgtgtcttttgttc	1080
QY	1081	atgcattggtctttgcgaagaaaaaagacaagaacaagaaaaaagacaagacagaga	1140
DB	1081	atgcattggtctttgcgaagaaaaaagacaagaacaagaaaaaagacaagacagaga	1140
QY	1141	caaaaacgaatcacacaacccaacttcaatagtcaatgctgcatcaagaatcgcgcgtcc	1200
DB	1141	caaaaacgaatcacacaacccaacttcaatagtcaatgctgcatcaagaatcgcgcgtcc	1200
QY	1201	atgtaatgtcttaaatgccaatgcagaagcaaacagtgcttaacatgtaactttaatgtctac	1260
DB	1201	atgtaatgtcttaaatgccaatgcagaagcaaacagtgcttaacatgtaactttaatgtctac	1260
QY	1261	ccatccaacccaacacaacacacatctgccttttccatcatcaacaacaacctgt	1320
DB	1261	ccatccaacccaacacaacacacatctgccttttccatcatcaacaacaacctgt	1320
QY	1321	atataatcatctcttcgcgcaccccaatttcttcaacttcaacacacagtcacactgcaat	1380
DB	1321	atataatcatctcttcgcgcaccccaatttcttcaacttcaacacacagtcacactgcaat	1380
QY	1381	tgcgtgtatcccatgtgcccaaatctccaatgcatggttccaacaccttctcttatataa	1440
DB	1381	tgcgtgtatcccatgtgcccaaatctccaatgcatggttccaacaccttctcttatataa	1440
QY	1441	tacctataaataactcctaataataatcaatcactcttcttcatatccaatccaagtgta	1500
DB	1441	tacctataaataactcctaataataatcaatcactcttcttcatatccaatccaagtgta	1500
QY	1501	ctactctactactataataccaccaaccaactcatatccaatactactactatgaact	1560
DB	1501	ctactctactactataataccaccaaccaactcatatccaatactactactatgaact	1560
QY	1561	tccttaagctcttcccttcttaagcttctccttggtttgggtcaaatacttgttgcgtta	1620
DB	1561	tccttaagctcttcccttcttaagcttctccttggtttgggtcaaatacttgttgcgtta	1620

OY 1621 ctcaagctgctgagatcacccgcgcatctctctatacaaaagttaagctctccgttaagcgc 1680
|||||
Db 1621 ctcaagctgctgagatcacccgcgcatctctctatacaaaagttaagctctccgttaagcgc 1680
OY 1661 tgaaggaaacatgagactcttaagaagactctctgcagaaacaacagtatgcatcagcagca 1740
|||||
Db 1661 tgaaggaaacatgagactcttaagaagactctctgcagaaacaacagtatgcatcagcagca 1740
OY 1741 agtactccggcttcggtgaagtctgtagcgctgcacacttacaacacttataatgaat 1800
|||||
Db 1741 agtactccggcttcggtgaagtctgtagcgctgcacacttacaacacttataatgaat 1800
OY 1801 accttgggaagaatctaacctcggaaaccccgccctcaagagtacacgcttctcttgaatg 1860
|||||
Db 1801 accttgggaagaatctaacctcggaaaccccgccctcaagagtacacgcttctcttgaatg 1860
OY 1861 gttcctctgactcttggtgtccctctatctactgcagaagcaatgctgcagaagaacac 1920
|||||
Db 1861 gttcctctgactcttggtgtccctctatctactgcagaagcaatgctgcagaagaacac 1920
OY 1921 aaagattcgatccgagaagaagctcgtccacacttccagaacttgggcaaacctgtctatac 1980
|||||
Db 1921 aaagattcgatccgagaagaagctcgtccacacttccagaacttgggcaaacctgtctatac 1980
OY 1981 actacggttacaggttagcatgcaagaaatcttaaggctatgataccgctacgtctccaca 2040
|||||
Db 1981 actacggttacaggttagcatgcaagaaatcttaaggctatgataccgctacgtctccaca 2040
OY 2041 ttggtgagatctcaacagacaaataagacttagcacccaaagaacaggtgtatcttcac 2100
|||||
Db 2041 ttggtgagatctcaacagacaaataagacttagcacccaaagaacaggtgtatcttcac 2100
OY 2101 atgacagaatctcgatcgagcctctgtgatgcatcccaacgcctgcgcagagtactcga 2160
|||||
Db 2101 atgacagaatctcgatcgagcctctgtgatgcatcccaacgcctgcgcagagtactcga 2160
OY 2161 taectgtgttgacaacaatgatacgcgaacactagtagctcaagaactgttctcggtt 2220
|||||
Db 2161 taectgtgttgacaacaatgatacgcgaacactagtagctcaagaactgttctcggtt 2220
OY 2221 acatgagacaggaatgagcagagagagcatgctcacagcttgaggctatgtatccact 2280
|||||
Db 2221 acatgagacaggaatgagcagagagagcatgctcacagcttgaggctatgtatccact 2280
OY 2281 acacaggaatctctcaactggttccagctactgtgcagcagtaactggaattcaactg 2340
|||||
Db 2281 acacaggaatctctcaactggttccagctactgtgcagcagtaactggaattcaactg 2340
OY 2341 acagtgtaacacatcagcgtgtgtgtgtgcatgtgaaagtgtaagctatctctg 2400
|||||
Db 2341 acagtgtaacacatcagcgtgtgtgtgtgcatgtgaaagtgtaagctatctctg 2400
OY 2401 ataccggttagctccaagctggttcggaactagcagcgacatctcacaactltagcaagcta 2460
|||||
Db 2401 ataccggttagctccaagctggttcggaactagcagcgacatctcacaactltagcaagcta 2460
OY 2461 ttggtgagacacacagaaacaggtgagtttgacatagatltgcgaacaacttgaactaca 2520
|||||
Db 2461 ttggtgagacacacagaaacaggtgagtttgacatagatltgcgaacaacttgaactaca 2520
OY 2521 tgcctacagttgtcttltgagatcaacgcaagatgtaccactgacccctccgctata 2580
|||||
Db 2521 tgcctacagttgtcttltgagatcaacgcaagatgtaccactgacccctccgctata 2580
OY 2581 ccagccacggaatcaaggttctgcgaacccagttgatactccagagtgagaacatctccagaat 2640
|||||
Db 2581 ccagccacggaatcaaggttctgcgaacccagttgatactccagagtgagaacatctccagaat 2640
OY 2641 ggaacttggagatgctgtcatctgtagtactacagagcttltgcagcgggccaacaac 2700
|||||
Db 2641 ggaacttggagatgctgtcatctgtagtactacagagcttltgcagcgggccaacaac 2700
OY 2701 tcgltgggctagctaaagcaatctgaagcttaataagtaagtaactaaatgcatgtagt 2760
|||||

|||||
Db 2701 tcgltgggctagctaaagcaatctgaagcttaataagtaagtaactaaatgcatgtagt 2760
OY 2761 gtaagagctcatgtagagacatggaatatgtatctgcgaacatgtaacagtaataactcga 2820
|||||
Db 2761 gtaagagctcatgtagagacatggaatatgtatctgcgaacatgtaacagtaataactcga 2820
OY 2821 gttccatccacttctctatgataaacaagaggtgtatgatatataactacactatc 2880
|||||
Db 2821 gttccatccacttctctatgataaacaagaggtgtatgatatataactacactatc 2880
OY 2881 tatgacacttattgttctcatgataaattccctctattatataataatcatctgaactg 2940
|||||
Db 2881 tatgacacttattgttctcatgataaattccctctattatataataatcatctgaactg 2940
OY 2941 acggcttatggaatgcttcaaatagtaacaaaacaatgtgtactataagaacttctaa 3000
|||||
Db 2941 acggcttatggaatgcttcaaatagtaacaaaacaatgtgtactataagaacttctaa 3000
OY 3001 caattctaactttagcatgtggaacgagacataaggtgttaagaaagacataacatataa 3060
|||||
Db 3001 caattctaactttagcatgtggaacgagacataaggtgttaagaaagacataacatataa 3060
OY 3061 tggaaagaagttgtctccattatataatataatataccactttagtataataga 3120
|||||
Db 3061 tggaaagaagttgtctccattatataatataatataccactttagtataataga 3120
OY 3121 tgttaagagacataacaatatataaagagaaggtgttatccattatataatataac 3180
|||||
Db 3121 tgttaagagacataacaatatataaagagaaggtgttatccattatataatataac 3180
OY 3181 taccattatataatataactataccacttattatgcttcttaagaagttgataccgt 3240
|||||
Db 3181 taccattatataatataactataccacttattatgcttcttaagaagttgataccgt 3240
OY 3241 atatttctaataatttagtgtatagatglatatgaaaggtactatttgaactcttactc 3300
|||||
Db 3241 atatttctaataatttagtgtatagatglatatgaaaggtactatttgaactcttactc 3300
OY 3301 tgtataaaggttggatcatcctcttaagaaggtgtctatattatatttgccttcaagat 3360
|||||
Db 3301 tgtataaaggttggatcatcctcttaagaaggtgtgtctatattatatttgccttcaagat 3360
OY 3361 aaaaaaaaatctatgagttgtgttggataaataatgaaaggttaaaaaataataa 3420
|||||
Db 3361 aaaaaaaaatctatgagttgtgttggataaataatgaaaggtttaaaaataataa 3420
OY 3421 ataaataacataataatataatglatataaattatataataataacattatctataa 3480
|||||
Db 3421 ataaataacataataatataatglatataaattatataataataacattatctataa 3480
OY 3481 agtaaatattgtcataaataatcataacaatcggtttagccttgcgtgaacgactcctaattc 3540
|||||
Db 3481 agtaaatattgtcataaataatcataacaatcggtttagccttgcgtgaacgactcctaattc 3540
OY 3541 taaacgagagtaaacataattggaactttagtattttaaacaatatattatataacact 3600
|||||
Db 3541 taaacgagagtaaacataattggaactttagtattttaaacaatatattatataacact 3600
OY 3601 atgaaattttttttttatccgcaaggaataaataatgaaatgaaagggacaatgtgt 3660
|||||
Db 3601 atgaaattttttttttatccgcaaggaataaataatgaaatgaaagggacaatgtgt 3660
OY 3661 gtcccaatccctatacaacaacactccacagaaggtgcaggtcggggacacaaaanaac 3720
|||||
Db 3661 gtcccaatccctatacaacaacactccacagaaggtgcaggtcggggacacaaaanaac 3720
OY 3721 aggcaaggaataatttttaatttgggtgtgtctgttgcgtcatatatttagcaggtaaa 3780
|||||
Db 3721 aggcaaggaataatttttaatttgggtgtgtctgttgcgtcatatatttagcaggtaaa 3780
OY 3781 caataacataacccctttagcagtagagcaatgtgtgacggtgtgctttagcttctt 3840
|||||

Db 3781 cactacataacccttagagagacatggtgaccgigtgcttagctcttta 3840
QY 3841 ttatttttttaccagaagaataaataatagacacttcagagatgttc 3900
Db 3841 ttatttttttaccagaagaataaataatagacacttcagagatgttc 3900
QY 3901 aacctatacaaaccccaaacagttctcctagacaccctacaactaagttacc 3957
Db 3901 aacctatacaaaccccaaacagttctcctagacaccctacaactaagttacc 3957
RESULT 2
AAZ10376
ID AAZ10376 standard; DNA: 1558 BP.
XX
AC AAZ10376;
XX
DT 15-NOV-1999 (first entry)
XX
DE Nucleotide sequence of the bean phaseolin promoter.
XX
KW Fao gene: plant metabolism; fatty acid oxidation enzyme;
KM polyhydroxyalkanoate; oil composition; seed production; plant biomass;
KW transgenic plant; promoter; bean; ss.
XX
OS Phaseolus sp.
XX
PN W09945122-A1.
XX
PD 10-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US04999.
XX
PR 06-MAR-1998; 98US-0077107.
XX
PA (META-) METABOLIX INC.
XX
PI Boynton L, Huisman GW, Moloney M, Patterson N, Peoples OP;
PI Snell K;
XX
DR WPI: 1999-540850/45.
XX
PT Modifying fatty acid metabolism in plants, useful for increasing
PC biomass and producing specific polymers in seeds
XX
PS Example 3; Page 63; 79pp; English.
XX
CC The present sequence represents the bean phaseolin promoter, which
CC is used to construct plasmids for the expression of the P. putida
CC faoAB gene. The gene encodes an enzyme that may be used in the
CC method of the invention. The specification describes a method for
CC manipulating the metabolism of a plant, and comprises expressing a
CC heterologous gene encoding fatty acid oxidation enzymes in the
CC cytosol or plastids other than the peroxisomes, glyoxisomes or
CC mitochondria of the plant. The method may be used to enhance the
CC biological production of polyhydroxyalkanoates or novel oil compositions
CC in a transgenic plant. Plants which may be used to produce these
CC compounds in this way include Brassicas, maize, soybean, cottonseed,
CC sunflower, palm, coconut, safflower, peanut, mustards, flax, tobacco and
CC alfalfa. The method may also be used to prevent or suppress seed
CC production and therefore increase the production of biomass (leaves,
CC stems, stalks) by plants.
XX
SQ Sequence 1558 BP; 516 A; 310 C; 171 G; 561 T; 0 other;

Query Match 39.0%; Score 1544; DB 20; Length 1558;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 catgtactccagatatacatatagtaagtttggtctctcgcggtgttttttac 71
Db 8 catgtactccagatatacatatagtaagtttggtctctcgcggtgttttttac 67

QY 72 ctctattaaagggtttccacctaataaatctcgtgatcatctcacttactgttac 131
Db 68 ctctattaaagggtttccacctaataaatctcgtgatcatctcacttactgttac 127
QY 132 ttaatttctcaataacttggttgtgaattatcaagcgtccgcacagataccctcaaa 191
Db 128 ttaatttctcaataacttggttgtgaattatcaagcgtccgcacagataccctcaaa 187
QY 192 atttattattgttaaacattttcaacccgataaattttatgaagttccgctactct 251
Db 188 atttattattgttaaacattttcaacccgataaattttatgaagttccgctactct 247
QY 252 taatgtagtcaacattttcatatgaataatataatattactaatttttagcgttggtag 311
Db 248 taatgtagtcaacattttcatatgaataatataatattactaatttttagcgttggtag 307
QY 312 aaagcataaagaattatcttctctctctccatataaagtttatatacataataaca 371
Db 308 aaagcataaagaattatcttctctctctccatataaagtttatatacataataaca 367
QY 372 aatctttaccttgaagaagattcccatattatattttaaaatatattacaatat 431
Db 368 aatctttaccttgaagaagattcccatattatattttaaaatatattacaatat 427
QY 432 ttccaacacgtaaatctcataataaagttgtttccaagaagataaaaatttaactcc 491
Db 428 ttccaacacgtaaatctcataataaagttgtttccaagaagataaaaatttaactcc 487
QY 492 ataattttttatctcgcagtactttaaagcaacacccagtgacaacactagcattttc 551
Db 488 ataattttttatctcgcagtactttaaagcaacacccagtgacaacactagcattttc 547
QY 552 tctttgaaataaaaatccaattatcatgtattttttttatcaaatgaattccacca 611
Db 548 tctttgaaataaaaatccaattatcatgtattttttttatcaaatgaattccacca 607
QY 612 aacaatcatcttggtgatcttctgaagcaagtcagttatgcaaaattctataattcccat 671
Db 608 aacaatcatcttggtgatcttctgaagcaagtcagttatgcaaaattctataattcccat 667
QY 672 ttgacacactggaagtaactgaagatctgcttttaacatgagagacactctctcaagta 731
Db 668 ttgacacactggaagtaactgaagatctgcttttaacatgagagacactctctcaagta 727
QY 732 attttaataagttacacatatccaagttccatatataatcaactacatactactctca 791
Db 728 attttaataagttacacatatccaagttccatatataatcaactacatactactctca 787
QY 792 aaaaattaatagatataataataatattactttttaatttaagttaagtgtgaat 851
Db 788 aaaaattaatagatataataataatattactttttaatttaagttaagtgtgaat 847
QY 852 ttgtagactatgatttatctactctagttaaatgttttatagatatttaagta 911
Db 848 ttgtagactatgatttatctactctagttaaatgttttatagatatttaagta 907
QY 912 aataaagtaagtgaagtaggttaagtggttttaacctcaaacataaactataacattta 971
Db 908 aataaagtaagtgaagtaggttaagtggttttaacctcaaacataaactataacattta 967
QY 972 ttgtgtagtaattttcatatatttcttatttgctttactctttcttgtagtaagtcgcg 1031
Db 968 ttgtgtagtaattttcatatatttcttatttgctttactctttcttgtagtaagtcgcg 1027
QY 1032 taactagaatacagttggttgccaatggcactcgtgtgcttttggttcacatgcatgggtc 1091
Db 1028 taactagaatacagttggttgccaatggcactcgtgtgcttttggttcacatgcatgggtc 1087
QY 1092 ttgcgcaagaaaagacaaaagaaaagaaaagaaaagaaaagaaaagaaaagaaaagcaat 1151
Db 1088 ttgcgcaagaaaagacaaaagaaaagaaaagaaaagaaaagaaaagaaaagcaat 1147

```
QY 1152 cacacaaccaactcaaatagtcactgctgtaacagatcgccgctgcattgta 1211
    |||
Db 1148 cacacaaccaactcaaatagtcactgctgtaacagatcgccgctgcattgta 1207
QY 1212 aatgcattgcaagaacacacgctgcttaataatgactttaatggtcaccctcaac 1271
    |||
Db 1208 aatgcattgcaagaacacacgctgcttaataatgactttaatggtcaccctcaac 1267
QY 1272 cacacaaccaactgctctttcttcatactcaacacacacactgtatatattcatt 1331
    |||
Db 1268 cacacaaccaactgctctttcttcatactcaacacacacactgtatatattcatt 1327
QY 1332 ctcttcgccacctcaattctcttcacttcaacacacgctcaactgcatatgctgcatc 1391
    |||
Db 1328 ctcttcgccacctcaattctcttcacttcaacacacgctcaactgcatatgctgcatc 1387
QY 1392 ccattgcccaaatctcattgcatgcttccaaacacctctctctataataataactataat 1451
    |||
Db 1388 ccattgcccaaatctcattgcatgcttccaaacacctctctctataataataactataat 1447
QY 1452 accttaatatcactcactcttcttcatactcatccatccagagtaactactactacta 1511
    |||
Db 1448 accttaatatcactcactcttcttcatactcatccatccagagtaactactactacta 1507
QY 1512 ctataatacccccaaccaactcatatactcaatactactactact 1555
    |||
Db 1508 ctataatacccccaaccaactcatatactcaatactactactact 1551
```

RESULT 3

AAx90961
AAx90961 standard; DNA; 1547 BP.

AC AAX90961;

DT 17-JAN-2000 (first entry)

DE Phaseolus vulgaris beta phaseolin gene 5' regulatory sequence.

KW Palmitate-CoA delta-9-desaturase; PCA: beta phaseolin gene; fatty acid;

KM palmitic acid; desaturase; modification; saturation; oil; transgenic plant;

KW 5' regulatory element; plasmid phagN184-2; tobacco; transformed plant; ss.

OS Phaseolus vulgaris.

XX Key location/Qualifiers

FT old_sequence replace(1049, C)

FT /note="Corresponds to base 1113 of plasmid phagN184-2

FT and facilitates subsequent cloning"

XX MO9950430-A2.

XX PD 07-OCT-1999.

XX PE 29-MAR-1999; 99MO-US06765.

XX PR 30-MAR-1998; 98US-0079840.

XX PA (DOWC) DOW AGROSCIENCES LLC.

XX PI Folkerets O, Merlo DJ;

XX DR WPI; 1999-610860/52.

XX New isolated palmitoyl-CoA delta-9-desaturase nucleic acids, used to

XX produce transgenic plants with altered levels of fatty acids

XX Example 2; Page 68-69; 78pp: English.

XX The present sequence is the 5' regulatory sequence of the beta phaseolin

CC gene from Phaseolus vulgaris. It is used in the construction of the

CC plasmid phagN184-2. This regulatory sequence is suitable for high level

CC and seed specific expression in tobacco. This plasmid can be used for
CC producing transgenic plants having altered saturated oil profiles and
CC decreased fatty acid levels in transformed plants.

XX Sequence 1547 BP, 513 A; 306 C; 170 G; 558 T; 0 other;

Query Match 36.4%; Score 1439; DB 20; Length 1547;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 12 catgtactccagatcatcattagtgaaagtcttgctctcgcggctgtttttac 71
    |||
Db 7 catgtactccagatcatcattagtgaaagtcttgctctcgcggctgtttttac 66
QY 72 cctatttaaggaggtttcccaactaaatctggtatcattccactcttggttac 131
    |||
Db 67 cctatttaaggaggtttcccaactaaatctggtatcattccactcttggttac 126
QY 132 tttaattctcataatcctgtgtgaattatcacgcttcgcacagatccctacaa 191
    |||
Db 127 tttaattctcataatcctgtgtgaattatcacgcttcgcacagatccctacaa 186
QY 192 attattattgtttaaacatttccaacccgcaaaaatttitaagagtcgcgtactct 251
    |||
Db 187 attattattgtttaaacatttccaacccgcaaaaatttitaagagtcgcgtactct 246
QY 252 taatgtagctctaacatttccatattgaaatatataattacttaatttagcgttag 311
    |||
Db 247 taatgtagctctaacatttccatattgaaatatataattacttaatttagcgttag 306
QY 312 aagcataagaattatcttcttcttcttcataataaagtttaatatatacaataa 371
    |||
Db 307 aagcataagaattatcttcttcttcttcataataaagtttaatatatacaataa 366
QY 372 aattcttaaccttaagaaagattcccaatttataatttaaaaatatatttcaaat 431
    |||
Db 367 aattcttaaccttaagaaagattcccaatttataatttaaaaatatatttcaaat 426
QY 432 ttccaacacgtaaatctcataataataaagtgtttccaaagtaataaattactcc 491
    |||
Db 427 ttccaacacgtaaatctcataataataaagtgtttccaaagtaataaattactcc 486
QY 492 ataatttttttttgcagcagatcttaagcaacaccagtgacacacactgcatctt 551
    |||
Db 487 ataatttttttttgcagcagatcttaagcaacaccagtgacacacactgcatctt 546
QY 552 tcttggataaaaaaatccaattatcatgtatttttttatacaaaatgaatccaca 611
    |||
Db 547 tcttggataaaaaaatccaattatcatgtatttttttatacaaaatgaatccaca 606
QY 612 aacaatcatttgggtattctcgaagcaagtcagtgtatgcanaattctataatccat 671
    |||
Db 607 aacaatcatttgggtattctcgaagcaagtcagtgtatgcanaattctataatccat 666
QY 672 ttgacactacggaagttaactgagaatcgcgttttatacagcgagacaacatcttcaagta 731
    |||
Db 667 ttgacactacggaagttaactgagaatcgcgttttatacagcgagacaacatcttcaagta 726
QY 732 attttaataatgattactatattcaagattcataatatcaaatcaatactacttcta 791
    |||
Db 727 attttaataatgattactatattcaagattcataatatcaaatcaatactacttcta 786
QY 792 aaaaattatagatataataataatattacttttataatttaagttaattgttgat 851
    |||
Db 787 aaaaattatagatataataataatattacttttataatttaagttaattgttgat 846
QY 852 ttgtacattgattattattactctactgatttaaatggtttatagatagtttaagta 911
    |||
Db 847 ttgtacattgattattattactctactgatttaaatggtttatagatagtttaagta 906
QY 912 aataaagtaatgtagtagtgtagtgtaggtttacccaataaacaataaataaatt 971
    |||
```

[illegible]

XX Modifying fatty acid metabolism in plants, useful for increasing
PT biomass and producing specific polymers in seeds
XX
XX
PS Example 3; page 71; 79pp; English.
XX
XX The present sequence represents the bean phaseolin terminator, which
CC is used to construct plasmids for the expression of the P. putida
CC fAOAB gene. The gene encodes an enzyme that may be used in the
CC method of the invention. The specification describes a method for
CC manipulating the metabolism of a plant, and comprises expressing a
CC heterologous gene encoding fatty acid oxidation enzymes in the
CC cytosol or plastids other than the peroxisomes, glyoxisomes or
CC mitochondria of the plant. The method may be used to enhance the
CC biological production of polyhydroxyalkanoates or novel oil compositions
CC in a transgenic plant. Plants which may be used to produce these
CC compounds in this way include Brassicas, maize, soybean, cottonseed,
CC sunflower, palm, coconut, safflower, peanut, mustards, flax, tobacco
CC alfalfa. The method may also be used to prevent or suppress seed
CC production and therefore increase the production of biomass (leaves,
CC stems, stalks) by plants.

Sequence 1244 BP; 456 A; 174 C; 178 G; 436 T; 0 other;

Query Match	31.0%	Score 1236;	DB 20;	Length 1244;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 126; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Qy	2732	aaagatgtagcaaaaatccatgtagaggtgaaagatctatggaagagcatggaattcgt	2791
Dd	19	ataatgatacgacaaaatgcagctgaaggtgtaagagctccatggaagagcatggaattcgt	78
Qy	2792	atccgacacatglaacagatataaataactgagctccatctcaattctctatgataaaca	2851
Dd	79	atccgacacatgtaacagatataaataactgagctccatctcaattctctatgataaaca	138
Qy	2852	aggatgtaataataataaacctctctatgacacctatgcttctatgataaattcc	2911
Dd	139	aggatgtaataataataaacctctctatgacacctatgcttctctatgataaattcc	198
Qy	2912	tcttatatlaataatcatcgtgaatcgttgacgagcttatggaatgcttcaaatgataca	2971
Dd	199	tcttatatlaataatcatcgtgaatcgttgacgagcttatggaatgcttcaaatgataca	258
Qy	2972	aacaaatgtafaactataagacttctcgaacaattccaacttgaagatgtggaagagca	3031
Dd	259	aacaaatgtafaactataagacttctcgaacaattccaacttgaagatgtggaagagca	318
Qy	3032	taagtgcttaagaagacataacaattataatgagaagaattgctccacattatataat	3091
Dd	319	taagtgcttaagaagacataacaattataatgagaagaattgctccacattatataat	378
Qy	3092	atataccaccttctgtaataataataagatggttaagaagacataacaattataagaagag	3151
Dd	379	atataccaccttctgtaataataataagatggttaagaagacataacaattataagaagag	438
Qy	3152	aagtttgtatccattatataataataactaccacattatataataataactccaatha	3211
Dd	439	aagtttgtatccattatataataataactaccacattatataataataactccaatha	498
Qy	3212	tttaatgcttataaagtttgatccatgtaatttccaaattatgattgattgatatat	3271
Dd	499	tttaatgcttataaagtttgatccatgtaatttccaaattatgattgattgatatat	558
Qy	3272	gaaagagtaataatgtaactctctaaactcgataaaaggttgatacatccttaagaatg	3331
Dd	559	gaaagagtaataatgtaactctctaaactcgataaaaggttgatacatccttaagaatg	618
Qy	3332	tctatttaatttatcttctcttaacagataaaaaaaattatgaattgattgtaataa	3391
Dd	619	tctatttaatttatcttctcttaacagataaaaaaaattatgaattgattgtaataa	678

[illegible][illegible]

```
Qy 2094 ttacacatgcagaatctcgatgcgcatccttgtagtgcgcatacccatcgctcgctcagag 2153
|||||
Db 541 ttccacatgcagaatctcgatgcgcatccttgtagtgcgcatacccatcgctcgctcagag 600
Qy 2154 taactgcacctggtgttgtaacaatgatgaacccgaacctagtagctcaagactgttc 2213
|||||
Db 601 ttaccgataacctgtgttgtaacaatgatgaacccgaacctagtagctcaagactgttc 660
Qy 2214 tcggtttacatgcagaagaatgcgcagagagacatgcctcaacgcttgagacattatcca 2273
|||||
Db 661 tcggtttacatgcagaagaatgcgcagagagacatgcctcaacgcttgagacattatcca 720
Qy 2274 tctactacacagaatctcttcaactggttccagttcaactgctgcagcagtagctgaattc 2333
|||||
Db 721 tctactacacagaatctcttcaactggttccagttcaactgctgcagcagtagctgaattc 780
Qy 2334 actgtgcagctgcacacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2393
|||||
Db 781 actgtgcagctgcacacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 840
Qy 2394 atcttgataccggttagctcaaaagctgtgtgcagctagcagcgacattctcaacattcag 2453
|||||
Db 841 atcttgataccggttagctcaaaagctgtgtgcagctagcagcgacattctcaacattcag 900
Qy 2454 caagctattgcagccacacagaacacagtaacgtagtgcagtagattgcagacacactt 2513
|||||
Db 901 caagctattgcagccacacagaacacagtaacgtagtgcagtagattgcagacacactt 960
Qy 2514 agctacatgcctacaagttgtcttgagatcaacggcagaagtgtaccacatccccctcc 2573
|||||
Db 961 agctacatgcctacaagttgtcttgagatcaacggcagaagtgtaccacatccccctcc 1020
Qy 2574 gctataccacgcagatcaacgggttctgcacacagttgattccagaagtgaacacattcc 2633
|||||
Db 1021 gctataccacgcagatcaacgggttctgcacacagttgattccagaagtgaacacattcc 1080
Qy 2634 cagaatgcagctctgcagagatgtgttcattcgttgagtagtactacagcgtcttgcagagcc 2693
|||||
Db 1081 cagaatgcagctctgcagagatgtgttcattcgttgagtagtactacagcgtcttgcagagcc 1140
Qy 2694 aacaacctcgttgggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 2726
|||||
Db 1141 aacaacctcgttgggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1173

RESULT 6
AADI7528
ID AADI7528 standard; DNA; 2970 BP.
XX
AC AADI7528;
XX
Df 10-DEC-2001 (first entry)
XX
De Kunitz soybean trypsin inhibitor (KSTI) encoding DNA.
XX
KW Soybean; allergen; transgenic plant; P34 protein; Gly m Bd 30K; Gly m 1;
KW soybean vacuolar protein; Gly m IA; Gly m IB; RGLY m3; Glycinin G1;
KW alablb; food; infant formula; animal feed; coating; salad oil; syrup;
KW spraying oil; roasting oil; frying oil; cracker; confectionery product;
KW snack food; topping; sauce; batter; breeding mixture; baking mix; dough;
KW Kunitz soybean trypsin inhibitor; KSTI; KTI3; ds.
XX
OS Glycine max.
XX
PN WO200168887-A2.
XX
PD 20-SEP-2001.
XX
PE 15-MAR-2001; 2001WO-US08254.
XX
PR 16-MAR-2000; 2000US-0189823.
XX
```

```
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
PI Jung R, Kinney AJ;
XX
DR WPI: 2001-582460/65.
XX
PT Recombinant expression construct to lower allergen (e.g. Gly m Bd 30K)
PT content of a soybean, comprises a nucleic acid fragment encoding the
PT allergen, useful for producing soybean plants which can be used to make
PT soybean products
XX
PS Claim 4; Page 48-49; 57pp; English.
XX
CC The patent discloses hypoallergenic transgenic soybeans and recombinant
CC expression constructs to lower soybean vacuolar protein, commonly known
CC as P34 (Gly m Bd 30K or Gly m 1) and other allergens such as Gly m IA,
CC Gly m IB, RGLY m3 or Glycinin G1 (alablb). The allergen content of the
CC soybean is reduced by sense suppression which is accomplished by using
CC the expression construct that comprises a nucleic acid fragment encoding
CC the allergen. The constructs are useful for producing hypoallergenic
CC transgenic soybean plants which can be used to make hypoallergenic
CC soybean products which can be used in a variety of food (e.g. infant
CC formulas) and animal feed applications. The oil made from seeds of the
CC hypoallergenic transgenic soybean plants can be used as ingredients,
CC as coatings, as salad oils, as spraying oils, as roasting oils, and
CC as frying oils. The foods in which the oil may be used include crackers
CC and snack foods, confectionery products, syrups and toppings, sauces,
CC batter and breeding mixtures, baking mixes and doughs. The present
CC sequence is a DNA encoding kunitz soybean trypsin inhibitor (KSTI
CC or KTI3), a minor soybean seed allergen.
CC
SQ Sequence 2970 BP; 1018 A; 547 C; 527 G; 878 T; 0 other;

Query Match 22.4%; Score 887; DB 22; Length 2970;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 887; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2735 aagtagaaccaaatgcatgtgtagtgaagagctcatgagacatggaattgtatc 2794
|||||
Db 1796 aagtagaaccaaatgcatgtgtagtgaagagctcatgagacatggaattgtatc 1855
Qy 2795 cgacatgtaacagbataataacgtgagctccatctcattctctatgaatacaaaagg 2854
|||||
Db 1856 cgacatgtaacagbataataacgtgagctccatctcattctctatgaatacaaaagg 1915
Qy 2855 atgttatgatataataaacctctatctatgcaccttatgtgtctatgataattcctct 2914
|||||
Db 1916 atgttatgatataataaacctctatctatgcaccttatgtgtctatgataattcctct 1975
Qy 2915 tattaataataatcatctgtaacgtcgttaacgctttagatgtccaatgatacaaaac 2974
|||||
Db 1976 tattaataataatcatctgtaacgtcgttaacgctttagatgtccaatgatacaaaac 2035
Qy 2975 aaatgttactataaagacttcttaacaactctaaactttagatcatgtgacagagacataa 3034
|||||
Db 2036 aaatgttactataaagacttcttaacaactctaaactttagatcatgtgacagagacataa 2095
Qy 3035 gtgttaagaaacataacaatatataatgaagaagttgtcttcatttatatatata 3094
|||||
Db 2096 gtgttaagaaacataacaatatataatgaagaagttgtcttcatttatatatata 2155
Qy 3095 ttaccacattatgatatattagatgtttaagagacatacaaatataaagagagag 3154
|||||
Db 2156 ttaccacattatgatatattagatgtttaagagacatacaaatataaagagagag 2215
Qy 3155 ttgtatcatttatatatataatactaacattatataatataatccactattt 3214
|||||
Db 2216 ttgtatcatttatatatataatactaacattatataatataatccactattt 2275
Qy 3215 aatgtcttataaggtttgataccatgatatttctaataattttagttgatatatgaa 3274
|||||
```

```
Db 2276 aatgctttaaaggcttgatccatgataatcttaataatttagtgatglatatgaa 2335
QY 3275 agggactatttgaaactctctactctgtataaaggttgatcatccttaagtggtct 3334
    |||||||
Db 2336 agggactatttgaaactctctactctgtataaaggttgatcatccttaagtggtct 2395
QY 3335 attaatcttctgtctcttcaacagataaaaaaaatataagattggttctgataaatat 3394
    |||||||
Db 2396 attaatcttctgtctcttcaacagataaaaaaaatataagattggttctgataaatat 2455
QY 3395 tgaagagatttaaaataaataaataaataaataaataaataaataaataaataaataa 3454
    |||||||
Db 2456 tgaagagatttaaaataaataaataaataaataaataaataaataaataaataaataa 2515
QY 3455 tataataaacattatctatataaaaaagtaaatgtatgcataacatacctacatcgctta 3514
    |||||||
Db 2516 tataataaacattatctatataaaaaagtaaatgtatgcataacatacctacatcgctta 2575
QY 3515 gccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3574
    |||||||
Db 2576 gccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2635
QY 3575 tttaacaataattatcttaacacataatgaaatttttttttttttttttttttttttc 3621
    |||||||
Db 2636 tttaacaataattatcttaacacataatgaaatttttttttttttttttttttttttc 2682
```

RESULT 7

```
AAAN30063
ID AAAN30063 standard; cDNA; 1098 BP.
XX
AC AAAN30063;
XX
DF 14-JUN-1992 (first entry)
XX
DE Sequence of prorennin cDNA in PCR 10001.
XX
KW Renin; renin; enzyme; protease; ss.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT CDS 1..1098
FT /*tag= a
XX
PN EP73029-A.
XX
PD 02-MAR-1983.
XX
PF 19-AUG-1982; 82EP-0107601.
XX
PR 24-AUG-1981; 81JP-0131631.
XX
PA (BEPP/) BEPPU T.
XX
PI Beppu T, Uozumi T, Nishimori K;
XX
DR WPI: 1983-22976K/10.
DR P-PSDB; AAP30603.
XX
PT Plasmid contg. calf pro:rennin DNA - and transformed
    microorganisms
XX
PS Example; Page 20-23; 32pp; English.
XX
CC The inventors claim recombinant plasmids contg. the cDNA of calf
    prorennin. Specified plasmids are PCR 10001 and PCR2001 (contg. the
    whole sequence plus the lac promoter region). Also new are
    CC microorganisms transformed with the plasmids, esp. E. coli CRI (ATCC
    CC 391710) contg. plasmid PCR2001.
XX
SQ Sequence 1098 BP; 252 A; 326 C; 301 G; 219 T; 0 other;
```

```
Query Match 1.2%; Score 49; DB 4; Length 1098;
Best Local Similarity 100.0%; Pred. No. 1.5e-11;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

RESULT 8

```
AAO04683
ID AAO04683 standard; DNA; 1098 BP.
XX
AC AAO04683;
XX
DT 05-OCT-1990 (first entry)
XX
DE Sequence encoding calf pro-rennin.
XX
KW Pro-rennin; ds.
XX
OS Bos taurus.
XX
PN JP02109984-A.
XX
PD 23-APR-1990.
XX
PF 01-JAN-1988; 88JP-0302176.
XX
PR 01-JAN-1988; 88JP-0302176.
XX
PA (BEPP/) BEPPU T.
XX
DR WPI: 1990-168358/22.
DR P-PSDB; AAR05080.
XX
PT Complex plasmid and microbe - contains calf pro-rennin cDNA.
XX
PS Disclosure: 32; 13pp; Japanese.
XX
CC Protein product may be expressed in E.coli expression system from
    CC plasmid pBR322.
XX
SQ Sequence 1098 BP; 252 A; 327 C; 300 G; 219 T; 0 other;
```

```
Query Match 1.2%; Score 49; DB 11; Length 1098;
Best Local Similarity 100.0%; Pred. No. 1.5e-11;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2209 tgtctcggtttacatgacaggaatggccagagagatgctcacgct 2257
    |||||||
Db 581 tgtctcggtttacatgacaggaatggccagagagatgctcacgct 629
```

RESULT 9

```
AAAN40295
ID AAAN40295 standard; mRNA; 1175 BP.
XX
AC AAAN40295;
XX
DT 04-FEB-1992 (first entry)
XX
DE Sequence encoding a polypeptide displaying milk clotting activity.
XX
KW Cheese-making; recombinant protein; rennet substitute; milk clot; ss.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT sig_peptide 21..69
FT /*tag= a
```


XX Recombinant DNA constructs for expressing protein in milk -
PT contg. specific mammary gland transcription control region and
PT signal sequence, providing high yield and easy prod. recovery
XX
PS Disclosure; Page 21; 41pp; German.
XX
CC The rennin gene was isolated using the probe represented in AAQ14775.
CC It was used as heterologous peptide/protein together with parts of the
CC alpha-S1-casein gene in the prodn. of DNA constructs. The heterologous
CC peptide or protein may also be human insulin-like growth factor I.
CC The constructs provide high yields of the protein with simple recovery
CC from the milk. Activation of the gene occurs only in the mammary gland.
CC See also AAQ14050, AAQ14774-77.
CC
XX
SQ Sequence 1210 BP; 278 A; 356 C; 324 G; 252 T; 0 other;

Query Match 1.2%; Score 49; DB 12; Length 1210;
Best Local Similarity 100.0%; Pred. No. 1.4e-11;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2209 tgtctcgtttacatgacaggaatgcccagagagatctcagct 2257
|||||
Db 643 tgtctcgtttacatgacaggaatgcccagagagatctcagct 691

RESULT 12
AAAT03006
ID AAAT03006 standard; DNA; 1240 BP.
XX
AC AAAT03006;
XX
DT 13-JUN-1996 (first entry)
XX
DE Chymosin open reading frame.
XX
KM al-3; albino mutant; light-regulated; Neurospora; bread mould;
KW heterologous gene; expression; control; chymosin; ss.
XX
OS Mammalian sp.
XX
PN WO9530739-A1.
XX
PD 16-NOV-1995.
XX
PF 09-MAY-1995; 95MO-US05716.
XX
PR 10-MAY-1994; 94US-0240372.
XX
PA (UYHA-) UNIV HAWAII.
XX
PI Kato EK, Stuart WD;
XX
PI Kato EK, Stuart WD;
XX
DR WPI; 1995-404108/51.
XX
PT Nucleic acid for expression of heterologous protein - contains
PT albino promoter for light induced expression in filamentous fungi
XX
PS Example 2; Fig 7; 26pp; English.
XX
CC The DNA is that of a mammalian gene (open reading frame) encoding
CC chymosin. The gene was placed in operable linkage with the al-3
CC promoter (see AAAT03005). The al-3 gene controls the production of
CC geranyl geranyl pyrophosphatase (GGPP) synthetase. GGPP is a precursor
CC for carotenoids and xanthophylls. It has been shown that exposure to
CC light increases the transcription level of GGPP synthetase 15-45 fold.
CC Light activates a number of genes in the common bread mould, Neurospora.
CC This can be used to regulate the expression of genes encoding
CC heterologous proteins, e.g. chymosin, in recombinant production systems.
CC Use of a light-regulated promoter is a simple and effective way to
CC control expression and allows timing to be adapted to the physiological
CC status of the host.

XX
SQ Sequence 1240 BP; 274 A; 374 C; 339 G; 253 T; 0 other;

Query Match 1.2%; Score 49; DB 16; Length 1240;
Best Local Similarity 100.0%; Pred. No. 1.4e-11;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2209 tgtctcgtttacatgacaggaatgcccagagagatctcagct 2257
|||||
Db 700 tgtctcgtttacatgacaggaatgcccagagagatctcagct 748

RESULT 13
AAN40055
ID AAN40055 standard; DNA; 1278 BP.
XX
AC AAN40055;
XX
DT 02-FEB-1992 (first entry)
XX
DE Sequence of prochymosin gene.
XX
KW Prochymosin expression vector; E.coli trp operon; chymosin; ss.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT CDS 3..1130
FT /*tag= a
FT 1245..1250
FT /*tag= b
XX
PN EP121775-A.
XX
PD 17-OCT-1984.
XX
PF 07-MAR-1984; 84EP-0102451.
XX
PR 09-MAR-1983; 83JP-0038439.
XX
PA (BEPF/) BEPPU T.
XX
PI Beppu T, Uozumi T, Nishimori K, Shimizu N, Kawaguchi Y;
PI Hidaka M;
XX
DR WPI; 1984-258001/42.
XX
DR P-PSDB; AAP40078.
XX
PT Expression plasmid comprising prochymosin gene and vector -
PT useful for transforming Escherichia coli for prochymosin prodn.
XX
PS Disclosure; Fig 1; 59pp; English.
XX
CC The inventors claim the prochymosin gene comprising a nucleotide
CC sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or
CC (b) the 5th codon (CGG) to the 365th codon (ATC); and recombinant
CC plasmids harboured by Escherichia coli strains deposited as FERM BP-
CC 262, -263 and -264. Any portion of the nucleotide sequence as
CC described in AAN40055 can be used. Also claimed is a vector derived
CC from plasmid pBR322. Typically plasmid pCR501 is obtd. from pOCT 2.
CC The transcriptional direction of pOCT 3 is opposite to that of
CC pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter
CC clockwise in pOCT 3.
XX
SQ Sequence 1278 BP; 309 A; 384 C; 326 G; 259 T; 0 other;

Query Match 1.2%; Score 49; DB 5; Length 1278;
Best Local Similarity 100.0%; Pred. No. 1.4e-11;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2209 tgtctcgtttacatgacaggaatgcccagagagatctcagct 2257

Db 613 tttctcggttacatgacaggaatgccaagagagacatgctcacgct 661

RESULT 14

AAAN30022
ID AAAN30022 standard; DNA; 1289 BP.

AC AAAN30022;

DT 25-APR-1992 (first entry)

DE Sequence of veal chymosin gene.

XX Protolytic enzyme; zymogen; rennin; chymosin; cheese making;
KM microbial vector; ss.

OS Bos taurus.

XX Key Location/Qualifiers

FT CDS 1..1140

XX BE897201-A.

PD 03-NOV-1983.

PF 30-JUN-1983; 83BE-0017731.

PR 13-APR-1983; 83US-0484539.

XX 01-JUL-1982; 82US-0394433.

PA (GENE-) GENEX CORP.

DR WPI: 1983-820813/47.

DR P-PSDB; AAP30013.

PT Isolated chymosin or rennin and prochymosin genes - plasmid(s)
PT which replicate in prokaryotic organisms, esp. Escherichia coli,
PT and organisms used for chymosin biosynthesis

XX Claim 7; Page 33-36; 43pp; French.

XX The inventors claim isolated chymosin (rennin) and prochymosin genes
CC from calves, and plasmids contg. the genes which are capable of
CC replicating in a prokaryotic organism. The prokaryotic organism is
CC pref. an Escherichia species, esp. E. coli p Gx 1225 (NRRL B-15061).

CC The microorganisms transformed by the plasmid are also claimed.

XX Sequence 1289 BP; 310 A; 391 C; 327 G; 260 T; 1 other;

SQ

Query Match 1.2%; Score 49; DB 4; Length 1289;

Best Local Similarity 100.0%; Pred. No. 1.4e-11; Indels 0; Gaps 0;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2209 tttctcggttacatgacaggaatgccaagagagacatgctcacgct 2257

Db 623 tttctcggttacatgacaggaatgccaagagagacatgctcacgct 671

AAAN91157
ID AAAN91157 standard; DNA; 1291 BP.

AC AAAN91157;

DT 07-JUN-1990 (first entry)

DE Cloned sequence of (pro)chymosin.

XX Lactic acid bacteria; cheese; Streptococcus cremoris SK112; chymosin;
KW proteinase; PSK112; ss.

XX Key Location/Qualifiers
FH precursor_RNA 29..1123
FT /*tag= a
FT /product=,prochymosin

XX NL8701378-A.

PD 02-JAN-1989.

PE 12-JUN-1987; 87NL-0001378.

XX 12-JUN-1987; 87NL-0001378.

PA (NEZU-) NEDERL. INS. ZUIVELON.

PI Simons AFM, De Vos WM;

DR WPI: 1989-030097/04.

DR P-PSDB; AAP94144.

PT DNA fragment having region specific for lactic acid bacteria -
PT is contained in plasmid in microorganism used in prodn. of
PT protein and food prodn. eg cheese.

XX Disclosure; fig 2; 43pp; Dutch.

XX The DNA encoding prochymosin can be cloned into a plasmid (esp. from
CC S. cremoris SK112) and used to produce large amts of the protein by
CC recombinant DNA techniques. This could overcome the shortage of
CC prochymosin due to shortage of calf stomachs and increasing cheese
CC prodn. Prochymosin is also used in prodn. of yoghurt, butter and
CC buttermilk. See also AAAN91158-N91160.

XX Sequence 1291 BP; 323 A; 385 C; 326 G; 257 T; 0 other;

SQ

Query Match 1.2%; Score 49; DB 10; Length 1291;

Best Local Similarity 100.0%; Pred. No. 1.4e-11;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2209 tttctcggttacatgacaggaatgccaagagagacatgctcacgct 2257

Db 609 tttctcggttacatgacaggaatgccaagagagacatgctcacgct 657

Search completed: August 1, 2002, 01:58:17

Job time: 16992 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 23:07:30 ; Search time 5849.81 Seconds
(without alignments)
9129.775 Million cell updates/sec

Title: US-09-643-755b-3

Perfect score: 3957
Sequence: 1 ctgcaggaatcatgtact.....accctaccactaaggtacc 3957

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estln:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	1.2	383	10	BG938086
2	49	1.2	430	10	BG937697
3	49	1.2	472	10	BG938320
4	43	1.1	399	10	BG937723
5	25	0.6	240	9	AU073125
6	25	0.6	668	9	AA394733
7	25	0.6	136	10	AG091550
8	24	0.6	214	10	BE763309
9	24	0.6	312	9	AU053870
10	24	0.6	442	9	AA441444
11	24	0.6	500	9	AU052887
12	24	0.6	560	12	BH366372
13	24	0.6	667	12	AO635511
14	24	0.6	702	10	BG872501
15	24	0.6	727	12	BH522720
16	24	0.6	782	10	BG432457
17					

18	24	0.6	892	10	BG399700	BG399700 602441568
19	23	0.6	178	10	BI974502	BI974502 sat169d07.
20	23	0.6	240	9	AU073229	AU073229 AU073229
21	23	0.6	298	9	BB493759	BB493759 BB493759
22	23	0.6	300	9	BB489847	BB489847 BB489847
23	23	0.6	304	12	AO907383	AO907383 GSSC0941
24	23	0.6	312	12	BH261085	BH261085 CH230-170
25	23	0.6	325	9	AV532492	AV532492 AV532492
26	23	0.6	339	10	T20903	T20903 2911 Lambda
27	23	0.6	380	9	AV538262	AV538262 AV538262
28	23	0.6	391	9	AA696042	AA696042 OP43d07.s
29	23	0.6	404	9	AA411567	AA411567 zv22g01.s
30	23	0.6	404	12	AO192651	AO192651 HS_2248_B
31	23	0.6	408	9	AV518232	AV518232 AV518232
32	23	0.6	411	9	AV559941	AV559941 AV559941
33	23	0.6	438	9	AA411566	AA411566 zv22g01.r
34	23	0.6	442	12	AO711285	AO711285 HS_5356_B
35	23	0.6	483	9	AV536021	AV536021 AV536021
36	23	0.6	505	9	AI479358	AI479358 tm27e07.x
37	23	0.6	511	9	AV518638	AV518638 AV518638
38	23	0.6	511	9	AV548208	AV548208 AV548208
39	23	0.6	532	9	AV557282	AV557282 AV557282
40	23	0.6	549	9	AV5107068	AV5107068 RPT-23-2
41	23	0.6	553	9	AV555646	AV555646 AV555646
42	23	0.6	564	12	A2457432	A2457432 1M0260F18
43	23	0.6	565	9	AV532027	AV532027 AV532027
44	23	0.6	574	9	AV542531	AV542531 AV542531
45	23	0.6	581	9	AV439574	AV439574 AV439574

ALIGNMENTS

RESULT 1

LOCUS BG938086 383 bp mRNA linear EST 11-JUN-2001
DEFINITION IAB01A08 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA sequence.
ACCESSION BG938086
VERSION BG938086.1 GI:14337458
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.
REFERENCE 1 (bases 1 to 383)
Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
CDNA's from bovine abomasum tissue
Unpublished (2001)
COMMENT Contact: Dr. Stephen Moore
. Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
The sequence best matches gb:BP019766 (Bos primigenius prothymosin mRNA, complete cds) in main database at high score of 735.0 and E-value of 0.0
PCR PRIMERS
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 383
POLYA-NO.

FEATURES

source
1..383
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bovine Abomasum cDNA Library"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"

```

/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMF"-strain"
/Note="Organ: Abomasum; Vector: Uni-22APXR; Site_1: EcoR
I; Site_2: Xho I"
BASE COUNT      83 a      111 c      113 g      76 t
ORIGIN

Query Match          1.2%; Score 49; DB 10; Length 383;
Best Local Similarity 100.0%; Pred. No. 2.7e-08;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  2209  tttctcggtttacatgacaggaatgccagagacatgtcagcgt 2257
|||||
Db  58  TGTCTCGGTTTACATGACAGGAATGCCAGAGACATGCTCAGCCT 106

RESULT  2
BG937697      430 bp  mRNA  linear  EST 11-JUN-2001
LOCUS      1AB005E01 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
DEFINITION      sequence.
ACCESSION      BG937697
VERSION        BG937697.1  GI:14337069
KEYWORDS
SOURCE
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 430)
Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
cDNA's from bovine abomasum tissue
Unpublished (2001)
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin)
mrna) in main database at high score of 844.0 and E-value of 0.0
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 430
POLYA-No.

FEATURES
source
location/Qualifiers
1..430
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bovine Abomasum cDNA Library"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMF"-strain"
/Note="Organ: Abomasum; Vector: Uni-22APXR; Site_1: EcoR
I; Site_2: Xho I"
BASE COUNT      97 a      130 c      119 g      84 t
ORIGIN

Query Match          1.2%; Score 49; DB 10; Length 430;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  2209  tttctcggtttacatgacaggaatgccagagacatgtcagcgt 2257
|||||
Db  52  TGTCTCGGTTTACATGACAGGAATGCCAGAGACATGCTCAGCCT 100
```

```

RESULT  3
BG938320      472 bp  mRNA  linear  EST 11-JUN-2001
LOCUS      1AB015E12 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
DEFINITION      sequence.
ACCESSION      BG938320
VERSION        BG938320.1  GI:14337692
KEYWORDS
SOURCE
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 472)
Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
cDNA's from bovine abomasum tissue
Unpublished (2001)
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin)
mrna) in main database at high score of 928.0 and E-value of 0.0
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 472
POLYA-No.

FEATURES
source
location/Qualifiers
1..472
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bovine Abomasum cDNA Library"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMF"-strain"
/Note="Organ: Abomasum; Vector: Uni-22APXR; Site_1: EcoR
I; Site_2: Xho I"
BASE COUNT      104 a      143 c      131 g      94 t
ORIGIN

Query Match          1.2%; Score 49; DB 10; Length 472;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  2209  tttctcggtttacatgacaggaatgccagagacatgtcagcgt 2257
|||||
Db  65  TGTCTCGGTTTACATGACAGGAATGCCAGAGACATGCTCAGCCT 113

RESULT  4
BG937723      399 bp  mRNA  linear  EST 11-JUN-2001
LOCUS      1AB005D06 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
DEFINITION      sequence.
ACCESSION      BG937723
VERSION        BG937723.1  GI:14337095
KEYWORDS
SOURCE
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 399)
```



```

RESULT 7
AG091550/c 660 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-091H03.R, genomic survey sequence.
DEFINITION AG091550
ACCESSION AG091550.1 GI:16643352
VERSION GSS: GSS (genome survey sequence).
KEYWORDS Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
SOURCE BAC library clone:PTB-091H03.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1 (bases 1 to 660)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
JOURNAL Unpublished
TITLE BAC end sequences of library PTB
REFERENCE 2 (bases 1 to 660)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
COMMENT Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbess@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pRS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1. 660
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-091H03.R"
/sex="male"
/cell_type="lymphoblast"
/clone_11b="PTB Chimpanzee Male BAC library"
BASE COUNT 128 a 166 c 130 g 236 t
ORIGIN
Query Match 0.6%; Score 25; DB 12; Length 660;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3403 ttaataataataataataata 3427
DB 94 TTAATAATAATAATAATAATAATA 70

```

```

RESULT 8
BE763309 136 bp mRNA linear EST 19-SEP-2000
LOCUS RC1.NT0033-090800-016-a01 NT0033 Homo sapiens CDNA, mRNA sequence.
DEFINITION BE763309
ACCESSION BE763309.1 GI:10193233
VERSION EST.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 136)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baie,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

```

```

TITLE M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
JOURNAL Shotgun sequencing of the human transcriptome with ORF expressed
MEDLINE sequence tags
COMMENT Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et3=RC1-NT0033-090
800-016-a01et3=2000-08-09et4=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 136.
Location/Qualifiers
1. 136
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="NT0033"
/dev_stage="Adult"
/note="Organ: nervous_tumor: Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 62 a 21 c 22 g 31 t
ORIGIN
Query Match 0.6%; Score 24; DB 10; Length 136;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3407 aataataataataataata 3430
DB 41 AATAATAATAATAATAATAATAACA 64

```

```

RESULT 9
C25503 214 bp mRNA linear EST 28-APR-1999
LOCUS C25503 Dictyostelium discoideum SL (H.Urushi-hara) Dictyostelium
DEFINITION C25503 Dictyostelium discoideum SLA138, mRNA sequence.
ACCESSION C25503
VERSION C25503.1 GI:2276053
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
REFERENCE 1 (bases 1 to 214)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mita,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
MEDLINE 99156227
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d40h@sakura.cc.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'

```

FEATURES POLYA-NO. Location/Qualifiers
SOURCE 1. .214
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLJ138"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 82 a 32 c 28 g 72 t
ORIGIN

Query Match 0.6%: Score 24; DB 10; Length 214;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3405 aaataataataataataata 3428
|||||
Db 160 AAATAATAATAATAATAATA 183

RESULT 10
AU053870 312 bp mRNA linear EST 28-APR-1999
DEFINITION AU053870 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
ACCESSION AU053870
VERSION AU053870.1 GI:4702352
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
1 (bases 1 to 312)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pl,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1998)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
Email: d402nuesakura.cc.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES Location/Qualifiers
SOURCE 1. .312
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLJ838"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 110 a 50 c 50 g 102 t
ORIGIN

Query Match 0.6%: Score 24; DB 9; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3405 aaataataataataataata 3428
|||||
Db 264 AAATAATAATAATAATAATA 287

RESULT 11
AA441444 442 bp mRNA linear EST 19-APR-2001
LOCUS LD16384.Sprime LD Drosophila melanogaster embryo Bluescript
DEFINITION Drosophila melanogaster cDNA clone LD16384 Sprime, mRNA sequence.
ACCESSION AA441444
VERSION AA441444.1 GI:4299185

KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 442)
AUTHORS Harvey,D., Brokslein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
TITLE BDGP/HHMT Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT On Jun 2, 1997 this sequence version replaced gi:2153322.
Contact: stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST/estefruitfly.berkeley.edu>
Plate: 163 row: G column: 12
High quality sequence stop: 254.
FEATURES Location/Qualifiers
SOURCE 1. .442
/organism="Drosophila melanogaster"
/db_xref="BDGP_EST:BDCL015599"
/db_xref="taxon:7227"
/clone="LD16384"
/clone_lib="LD Drosophila melanogaster embryo Bluescript"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="SOLR"
/note="Organ: embryo; Vector: Bluescript SK; Site: 1: SCORI
; Site: 2: xhoI; Constructed using Stratagene Zap-cDNA
Synthesis kit. Oligo dt-primed and directionally cloned at
EcoRI and xhoI in Bluescript SK(+/-)"
BASE COUNT 176 a 89 c 94 g 83 t
ORIGIN

Query Match 0.6%: Score 24; DB 9; Length 442;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3404 taaataataataataataata 3427
|||||
Db 165 TAAATAATAATAATAATAATA 188

RESULT 12
AU052887 500 bp mRNA linear EST 28-APR-1999
LOCUS AU052887 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
DEFINITION AU052887 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
ACCESSION AU052887
VERSION AU052887.1 GI:4701370
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
1 (bases 1 to 500)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pl,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1998)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
Email: d402nuesakura.cc.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES Location/Qualifiers
SOURCE 1. .500
/organism="Dictyostelium discoideum"

/strain="Ax4"
/db_xref="taxon:44689"
/clone="SLP272"
/clone_1lb="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 169 a 88 c 83 g 160 t
ORIGIN

Query Match 0.6%; Score 24; DB 9; Length 500;
Best Local Similarity 100.0%; Pred. No. 1e+02; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;

OY 3405 aaaaataataataataa 3428
|||||
DB 449 AAAATAATATAATATAATA 472

RESULT 13
LOCUS BH366372 560 bp DNA linear GSS 03-DEC-2001
DEFINITION CH230-204G22.T1 CHORI-230 Segment 1 Rattus norvegicus genomic clone
ACCESSION BH366372
VERSION BH366372.1 GI:17297106
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 560)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
A., Gebregorjts,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Frisner,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
Other_GSSs: CH230-204G22.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pjejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_eirng_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 204 row: G column: 22
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..560
/organism="Rattus norvegicus"
/strain="BN/SENHsd/MCM"
/db_xref="taxon:10116"
/clone="CH230-204G22"
/clone_1lb="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARBAC2.1; Site.1: EcoRI; Site.2: EcoRI;
CHORI-230 Rat (BN/SENHsd/MCM) BAC library produced by
Pieter de Jong"

BASE COUNT 195 a 103 c 89 g 173 t
ORIGIN

Query Match 0.6%; Score 24; DB 12; Length 560;
Best Local Similarity 100.0%; Pred. No. 98; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;

OY 398 catttattttaaataataact 421
|||||
DB 311 CATTATTATTATAATAATATT 334

RESULT 14
LOCUS AO635511/c 667 bp DNA linear GSS 17-JUN-1999
DEFINITION RPCI-11-490J20.TV RPCI-11 Homo sapiens genomic clone RPCI-11-490J20
DNA sequence.
ACCESSION AO635511
VERSION AO635511.1 GI:5098146
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 667)
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
J.C.
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
Unpublished (1997)
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..667
/organism="Homo sapiens"
/db_xref="GDB:7688011"
/db_xref="taxon:9606"
/clone="RPCI-11-490J20"
/clone_1lb="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI;
RPCI11 Human Male BAC library"

BASE COUNT 215 a 135 c 146 g 171 t
ORIGIN

Query Match 0.6%; Score 24; DB 12; Length 667;
Best Local Similarity 100.0%; Pred. No. 88; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;

OY 3405 aaaaataataataataa 3428
|||||
DB 288 AAAATAATATAATATAATA 265

RESULT 15
LOCUS BG872501 702 bp mRNA linear EST 29-MAY-2001
DEFINITION 602793535F1 NCL_CGAP_SG2 Mus musculus cDNA IMAGE:4924784 5',
mRNA sequence.
ACCESSION BG872501
VERSION BG872501.1 GI:14223041
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 702)
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM10848 row: e column: 09
High quality sequence stop: 655.
Location/Qualifiers

FEATURES
Source
1. 702
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4924784"
/clone_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: PCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT
ORIGIN
220 a 149 c 152 g 181 t

Query Match 0.6%; Score 24; DB 10; Length 702;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1109 aaagacaagaagaaagacaaga 1132
|||||
Db 666 AAAGACAAGAAAAAGACAAAA 689

Search completed: July 31, 2002, 23:07:38
Job time: 17283 sec

THIS PAGE BLANK (USPTO)